The role of lupus susceptibility MHC haplotypes and interferon-alpha in gene regulation

Salgado Teixeira Duarte, Carolina

Awarding institution:
King's College London

The role of lupus susceptibility MHC haplotypes and interferon-alpha in gene regulation

Carolina Salgado Teixeira Duarte

Submitted for the degree of Doctor of Philosophy (PhD)

Student number 1251472

Department of Genetics and Molecular Medicine

King's College London, 2016

Abstract

Systemic lupus erythematosus (SLE) is an autoimmune disease characterized by B cell dysregulation and type 1 interferon (IFN) activity. Variants in the major histocompatibility complex (MHC) region confer the greatest genetic risk for SLE; however, the underlying causes of association remain elusive. One mechanism by which causal variants may drive genetic associations at the MHC is through regulating gene expression in a context-specific manner. This project investigates the effect of (i) SLE-associated MHC haplotypes, and (ii) IFN-α stimulation on gene regulation in *ex vivo* B cells, in order to further our understanding of how these factors contribute to disease risk.

Expression quantitative trait loci (eQTLs) were investigated for variants tagging six SLE-risk or -protective haplotypes: *DRB1*03:01* (rs2187668), *DRB1*15:01* (rs3135391), *DPB1* (rs3117213, rs2071351, rs2071349), and *MSH5* (rs409558). eQTL analyses were conducted using publicly-available data sets. Additionally, gene expression data were generated from resting and IFN-α-stimulated *ex vivo* B cells. Several *cis*-eQTLs were identified and replicated in the publicly-available data sets. A novel *trans*-eQTL was identified for *DRB1*03:01* haplotypes in the exon array data set, in both resting and IFN-α-stimulated cells, involving down-regulation of *BTN3A2* ($P < 2.63 \times 10^{-6}$). These results suggest a regulatory role for disease-associated MHC haplotypes, and implicate *BTN3A2* as a novel candidate gene on the *DRB1*03:01* haplotype.

The global effect of IFN-α stimulation on the B cell transcriptome was also explored. Several SLE susceptibility genes outside canonical type 1 IFN signalling pathways were differentially expressed in response to IFN-α. The direction of effect of IFN-α on the expression of SLE candidate genes paralleled the known functional consequences of SLE-associated polymorphisms in those genes. This suggests a previously unrecognised role for SLE candidate genes following activation of the type 1 IFN pathway, and sheds light into the role of IFN-α in the aetiopathogenesis of SLE.

Table of Contents

Table of Figures

Table of Tables

Acknowledgements

I want to thank all colleagues who contributed to the research presented in this thesis. Firstly, my primary supervisor Dr Michelle Fernando for her thorough guidance and help, and my supervisor Professor Tim Vyse for his valuable insights. Secondly, the members of the Immunogenetics group and collaborators who assisted me whenever I sought their help – in particular, Dr David Morris, Dr James Bentham, Dr Andrea Cortini, Dr Amy Roberts, Ms Lingyan Chen, Mr Christopher Odhams, and Dr Matt Arno. I also want to thank Ms Elizabeth Manners for kindly agreeing to proofread this thesis.

My sincere thanks to Arthritis Research UK for funding this PhD studentship through programme grant 19983. Finally, I want to express my gratitude to my family for their constant support during this research.

Statement of Contribution

This thesis is the result of my own work, unless mentioned otherwise. Some of the data collection and analysis were performed in collaboration with colleagues, as follows:

The *trans*-eQTL data for the MuTHER LCL data set discussed in Chapter 3 were obtained from Dr Kerrin Small (King's College London).

Dr Michelle Fernando (King's College London) contributed to the conceptualization of the Affymetrix Human Exon 1.0 ST array gene expression study discussed in Chapters 4 and 5. Whole blood samples were provided by the TwinsUK adult resource at St. Thomas' Hospital. Blood samples were extracted by Mr Akmal Droubi, Dr Michelle Fernando, and Dr Myles Lewis (King's College London). The processing of RNA samples from 15 individual twins, and from all seven LCLs, was conducted by Ms Lora Boteva under the supervision of Dr Michelle Fernando (King's College London). cDNA synthesis and array hybridization were performed by Mr Erick Nasser, Dr Fei Wong, and Dr Estibaliz Aldecoa-Otalora Astarloa under the supervision of Dr Matt Arno at the Genomics Centre, King's College London. Mapping of SNP and indel-containing probes from the Affymetrix Human Exon 1.0 ST array conducted for quality control purposes was performed by Dr Mina Ryten and Dr Adaikalavan Ramasamy (King's College London).

Four-digit HLA typing of DNA samples was performed by Dr Robert Collins at the Clinical Transplantation Laboratory in Guy's Hospital. HLA imputation of TwinsUK samples was performed by Dr David Morris (King's College London).

Abbreviations

ANA	anti-nuclear antibodies
ANOVA	analysis of variance
APCs	antigen-presenting cells
BCR	B cell receptor
Be1	B effector 1
BSA	bovine serum albumin
BTN3	butyrophilin 3
cDNA	complementary DNA
ChIP-seq	chromatin immunoprecipitation-sequencing
CNV	copy number variant
Ct	threshold cycle
DABG	detection above background
DC	dendritic cell
DE	differentially expressed
DNA	deoxyribonucleic acid
dsDNA	double-stranded DNA
EBV	Epstein-Barr virus
eQTL	expression quantitative trait loci
eSNP	expression SNP
FC	fold change
FCGR	Fcγ receptor
FDR	false discovery rate
GoF	gain of function
GWAS	genome-wide association study
h	hour
HLA	human leukocyte antigen
IBD	inflammatory bowel disease
IC	immune complex

IFN	interferon
IFNAR	IFN-α receptor
IFNGR	IFN-γ receptor
IIM	idiopathic inflammatory myopathies
IPA	Ingenuity Pathway Analysis
IRF	IFN regulatory factor
ISRE	IFN-stimulated response element
KB	kilobase
kDa	kilodalton
LCL	lymphoblastoid cell line
LD	linkage disequilibrium
LoF	loss of function
LPS	lipopolysaccharide
MAF	minor allele frequency
MB	megabase
mDC	myeloid dendritic cell
MHC	major histocompatability complex
min	minute
miRNA	microRNA
mRNA	messenger RNA
MuTHER	Multiple Tissue Human Expression Resource
N	sample size
ND	no data available
NDE	non-differentially expressed
NFκB	nuclear factor κB
NHGRI	National Human Genome Research Institute
NK	natural killer
OR	odds ratio
Partek GS	Partek Genomics Suite
PBMC	peripheral blood mononuclear cell

PBS	phosphate-buffered saline
PC	principal component
PCA	principal component analysis
pDC	plasmacytoid DC
QC	quality control
qPCR	quantitative reverse-transcription polymerase chain reaction
reQTL	response eQTL
RNA	ribonucleic acid
RNA-seq	RNA sequencing
RRMS	relapsing remitting multiple sclerosis
RT	room temperature
sec	second
SLE	systemic lupus erythematosus
snoRNA	small nucleolar RNA
SNP	single nucelotide polymorphism
TCR	T cell receptor
TFBS	transcription factor binding site
Th1	T helper type 1
TLR	Toll-like receptor
TNF	tumour necrosis factor
Txn Factor	transcription factor
U.K.	United Kingdom
UTR	untranslated region
UV	ultraviolet
xMHC	extended MHC

Chapter 1 - Introduction

1.1 Systemic lupus erythematosus: a clinical overview

Systemic lupus erythematosus (SLE) is a chronic autoimmune disease that can affect multiple organ systems in the human body. The disease is characterized by the presence of autoantibodies against cellular and nuclear components, and tissue inflammation which can widely vary in severity. Despite the considerable advances in our understanding of disease pathogenesis achieved over the last few decades of research, the exact causes of SLE remain unknown.

1.1.1 Clinical presentation

SLE is a clinically heterogeneous disease in which patients can present a broad spectrum of symptoms. The manifestations of disease can range from mild to severe, typically including periods of relapse and remission. The most commonly reported symptoms include fatigue, joint pain (arthralgia), fevers, skin rashes, and photosensitivity. Some patients may also experience neuropsychiatric symptoms, such as seizures, strokes, migraines, and depression. More serious complications can arise from inflammation of vital organs, such as the kidneys (nephritis), the brain, and the heart (carditis), which can ultimately result in organ failure if left untreated (Lahita, 2004, Lisnevskaia et al., 2014, Mohan and Putterman, 2015). SLE patients exhibit antinuclear antibodies (ANA), and may show reduced numbers of peripheral blood cells and reduced levels of early complement components.

The presence of specific clinical manifestations (or sub-phenotypes) and autoantibody subsets has led many to define SLE as a spectrum of clinical syndromes, rather than a single disease. Lupus sub-phenotypes include the presence of renal disease (nephritis), skin involvement, and neurological involvement. Common subsets of ANA found in SLE include anti-double-stranded deoxyribonucleic acid (anti-dsDNA) antibodies, and anti-ribonuclear protein antibodies such as anti-Ro/SSA, anti-La/SSB, and anti-Sm (Lahita, 2004, Lisnevskaia et al., 2014). Approximately one third of SLE patients develop secondary autoimmune diseases – for example, thyroid disease, Sjögren's syndrome, antiphospholipid syndrome, or rheumatoid arthritis (Lockshin et al., 2015). As symptoms of SLE may be unspecific, diagnosis can be based on the presence of

at least four of eleven clinical features listed in the American College of Rheumatology classification criteria (Hochberg, 1997).

1.1.2 Epidemiology

SLE has a prevalence of 1 in 2,500 individuals in the United Kingdom (U.K.), and up to 150 in 100,000 individuals worldwide (O'Neill and Cervera, 2010, Tsokos, 2011). Disease incidence varies depending on ethnicity and geographical location, affecting more commonly non-European and migrant populations (namely African-American, Afro-Caribbean, and Hispanic-American). SLE mainly affects women of childbearing age, and may remit in patients over the age of 50. There is a strong gender bias towards females, with a nine to one ratio of female to male patients. SLE mortality rates have improved substantially in the last 60 years, with over 85% of patients currently surviving to 10 years. However, morbidity remains a significant problem due to the lack of effective therapy, damage caused by disease, and side effects from currently administered drugs (O'Neill and Cervera, 2010).

1.1.3 Aetiology

SLE is a complex disease, whereby multiple genetic, environmental, and hormonal factors together contribute to disease pathogenesis. Even though the cause of SLE is unknown, some insight into disease triggers has been gained. Probable environmental triggers include exposure to ultraviolet (UV) light, viral infections (namely Epstein-Barr virus), and drug administration (such as hydralazine and type 1 interferon, Gourley and Miller, 2007, O'Neill and Cervera, 2010, Jeffries and Sawalha, 2011). Hormonal factors are also thought to act as major disease triggers, and may explain the marked sex and age bias in disease incidence. In particular, oestrogen has been shown to enhance autoimmunity, and to trigger SLE relapse (O'Neill and Cervera, 2010). The higher incidence of SLE in women may also be caused by an X-chromosome dosage effect, as suggested by studies reporting increased disease rates in males with Klinefelter's syndrome (47, XXY karyotype), compared to males with a 46, XY karyotype (Dillon et al., 2011). Recent advances in the genetics of SLE susceptibility are described in section 1.4.

1.1.4 Current treatments

SLE is currently incurable. Treatments administered to patients attempt to reduce inflammation and manage symptoms, and consist mainly of immunosuppressants. The most common

therapeutic agents in use are non-steroidal anti-inflammatory drugs (NSAIDs), corticosteroids, anti-malarials (such as hydroxychloroquine), and biological agents (monoclonal antibodies). Therapeutic monoclonal antibodies currently in use target B cell-modulating components, such as the B cell-activating factor CD20 (targeted by Rituximab), or the B cell-survival factor BAFF (targeted by Belimumab). With the exception of Belimumab, no new drugs have been approved for the treatment of SLE in the last 50 years (Amissah-Arthur and Gordon, 2010). Furthermore, current treatments are not effective in all patients, and can cause a range of side effects, including infections and increased risk of cardiovascular disease. More research on the pathophysiology of SLE is needed to enable the development of targeted treatment for patients who do not respond well to conventional therapies, and for the identification of biomarkers of prognostic and diagnostic value (Liu and Davidson, 2012).

1.2 Immune dysregulation in SLE

Autoimmunity arises when an aberrant immune response to self-antigens is initiated, following a breakdown in immune tolerance. In healthy individuals, central and peripheral tolerance mechanisms ensure the deletion or inactivation of autoreactive B cells and T cells. In the thymus, thymic epithelial cells present self-antigens on major histocompatibility complex (MHC, or human leukocyte antigen) molecules to developing T cells, and signal autoreactive T cells to undergo apoptosis or become inactivated. Similarly, immature B cells encounter self-antigens in the bone marrow, where autoreactive B cells undergo apoptosis, inactivation, or B cell receptor (BCR) editing (central tolerance). At the periphery, autoreactive T cells become unresponsive following an antigen encounter in the absence of co-stimulatory signals (T cell anergy). Without CD4+ T cell help, autoreactive B cells cannot become activated (peripheral tolerance). In autoimmune disease, failure in tolerance mechanisms results in activation of autoreactive T cells and B cells, and the production of pathogenic autoantibodies (Liu and Davidson, 2012).

SLE is characterized by cellular dysfunction in both the innate and adaptive arms of the immune system. Three key pathogenic mechanisms in SLE include impaired clearance of apoptotic cells (the “waste disposal” hypothesis), increased type 1 interferon (IFN) activity (innate immunity), and B cell autoreactivity (adaptive immunity). The waste disposal hypothesis posits that

defective clearance of apoptotic cells leads to the release and accumulation of nucleic acid-containing self-antigens in the periphery. Presentation of self-nuclear antigens by antigen-presenting cells (APCs) in the presence of co-stimulatory signals subsequently leads to the activation of autoreactive B and T cells (Kelley et al., 2010). Environmental risk factors such as UV rays and viral infections are also thought to trigger increased cell death in SLE, thus precipitating the release of self-antigens. Both nuclear antigens and environmental triggers can induce the production of type 1 IFNs, namely IFN-α, by plasmacytoid dendritic cells (pDCs). The pro-inflammatory activity of IFN-α promotes maturation of APCs, such as myeloid dendritic cells (mDCs). mDCs present self-antigens in the context of human leukocyte antigen (HLA) molecules to T cells, thereby initiating adaptive immunity. CD4+ helper T cells release signals that promote B cell activation and differentiation into autoantibody-producing cells (plasma cells). Autoantibodies and self-antigens form immune complexes (ICs), which act as endogenous inducers of IFN-α production, thus perpetuating a vicious cycle of inflammation. Impaired clearance of cellular debris results in deposition of ICs in tissues, which can cause irreversible tissue and organ damage over time (Figure 1-1, Ronnblom, 2011, Liu and Davidson, 2012).

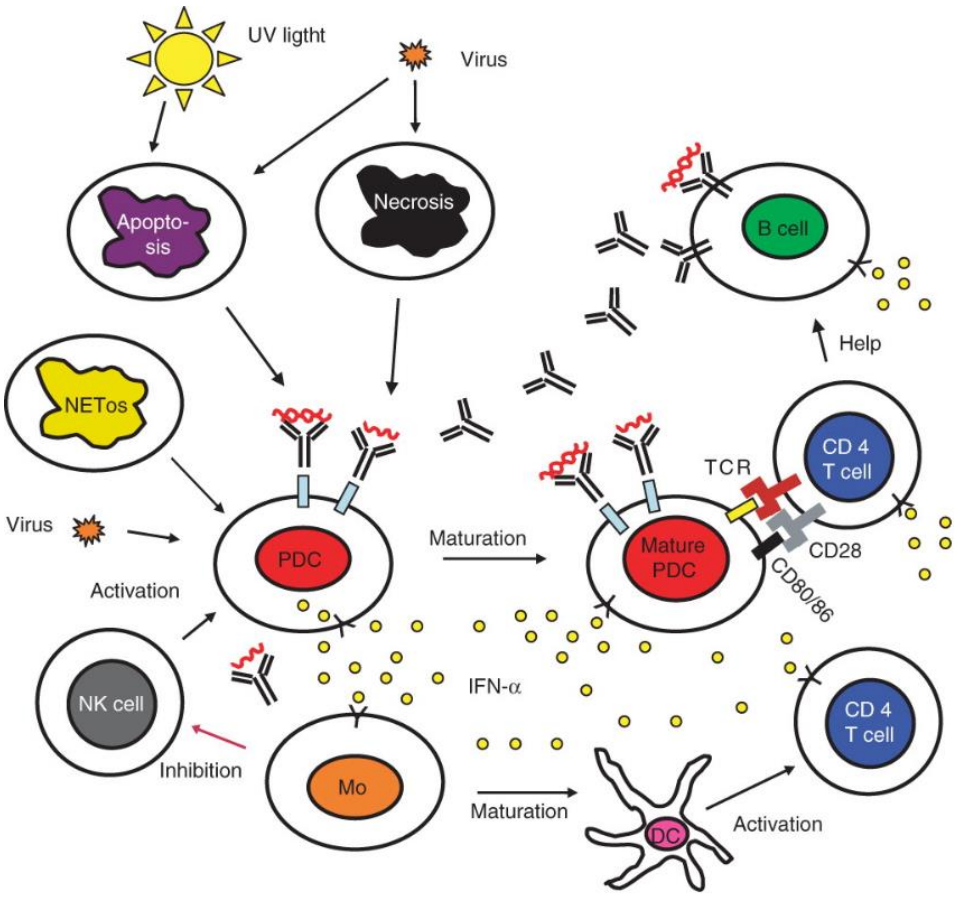


Figure 1-1 Illustration of the immunopathogenesis of SLE.

Environmental factors, such as viruses, trigger IFN-α production by plasmacytoid dendritic cells (PDC), and the release of self-antigens from apoptotic cells. IFN-α production induces maturation of dendritic cells, activation of T cells and B cells, and autoantibody production. Immune complexes (ICs) further induce IFN-α production, perpetuating a vicious cycle of inflammation. See text for a detailed description. NETosis – neutrophil cell death, NK cell – natural killer cell, Mo – monocyte, DC – dendritic cell, TCR – T cell receptor. Figure exported from Ronnblom (2011).

In the innate arm of the immune system, phagocytic cells such as neutrophils and macrophages play key roles in the clearance of cellular debris and ICs. For example, apoptotic cells are cleared by cross-binding to soluble phagocyte receptor ligands (such as complement factors) which induce phagocytosis. Similarly, Fcγ receptors (FCGRs) in phagocytic cells can induce phagocytosis of ICs by binding to their Fc portion, thus preventing IC accumulation and deposition in tissues (Pittoni and Valesini, 2002, Ronnblom, 2011). In SLE, impaired function of complement receptors and FCGRs in macrophages and/or neutrophils is thought to decrease clearance of ICs and apoptotic cells (Rhodes et al., 2012, Zhou et al., 2013).

1.3 The type 1 IFN pathway and SLE

A role for IFNs in SLE was first established in 1979, when IFN titres were found to be elevated in the serum of SLE patients (Hooks et al., 1979). Since then, there has been a considerable amount of evidence demonstrating the central role of type 1 IFNs, particularly IFN-α, in the pathogenesis of SLE. Sixty to eighty per cent of SLE patients demonstrate a gene expression profile associated with a response to type 1 IFN (i.e. a type 1 IFN signature) in multiple cell types, including B cells, T cells, and myeloid cells, as well as in kidneys (Baechler et al., 2003, Becker et al., 2013, Berthier et al., 2012). Elevated type 1 IFN activity has been shown in the serum of SLE patients, and correlates with disease activity and disease sub-phenotypes (Preble et al., 1982, Kirou et al., 2005, Bauer et al., 2006). Additionally, the administration of therapeutic IFN-α in chronic viral hepatitis and cancer patients can trigger autoimmunity, including lupus-like symptoms, in some patients (Ronnblom et al., 1991, Niewold and Swedler, 2005). Furthermore, exogenous IFN-α has been shown to accelerate the onset of SLE in a mouse model (Liu et al., 2011). The type 1 IFN signature has also been found in subsets of patients with other autoimmune diseases, including Sjögren's syndrome, rheumatoid arthritis, type 1 diabetes, systemic sclerosis, and idiopathic inflammatory myopathies, highlighting the importance of IFNs in autoimmunity (Mavragani et al., 2007, Mavragani et al., 2010, Crow, 2010, Assassi et al., 2010, Lundberg and Helmers, 2010).

1.3.1 The IFN family

IFNs are pro-inflammatory cytokines of the innate immune system which play a central role in anti-viral defence. IFNs have a variety of other functions, including anti-tumour response, the conferring of immunity against bacterial and fungal parasites, and fine-tuning of immune responses. Human IFNs have been divided into three groups: types 1, 2, and 3. Type 1 IFNs include 13 IFN-α subtypes, and the single members IFN-β, IFN-ε, IFN-κ, and IFN-ω. All type 1 IFNs share one heterodimeric receptor, the IFN-α receptor (IFNAR), which is composed of the IFNAR1 and IFNAR2 subunits. Differential binding of type 1 IFN subtypes to the IFNAR specifies the signalling cascades, and induces specific IFN signatures. Type 1 IFNs can be expressed by virtually any cell type, and share a degree of functional overlap with one another. This family of IFNs plays a central role in anti-viral defence and in the modulation of immune

responses. Type 2 IFN comprises a single member, IFN-γ, which signals through the IFN-γ receptor (IFNGR). IFN-γ is mostly expressed by natural killer (NK) cells in the innate response, and by activated T cells in the adaptive response. Type 3 IFNs comprise four members of the IFN-λ family, which are involved in anti-viral responses at mucosal surfaces. The redundancy of the IFN system ensures defence against pathogens, and highlights the importance of these cytokines in immunity (Ronnblom, 2011, Hoffmann et al., 2015).

1.3.2 Type 1 and type 2 IFN signalling

Depending on the type of antigen, type 1 IFNs can be produced via Toll-like receptor (TLR)-dependent or TLR-independent pathways. The production of type 1 IFNs can be initiated by the binding of nucleic acid-containing ICs to FCGRs, which triggers endocytosis of nuclear antigens. dsDNA and ribonucleic acid (RNA) antigens bind to endosomal TLR7 or TLR9 receptors, which activate a signalling cascade through MyD88, IRAK, and TRAF kinases, and IFN regulatory factors (IRFs) 3 and 7, which are transcriptional activators of IFN-α and IFN-β genes. Alternatively, cytoplasmic RNA activates the RIG-I and MDA5 cytoplasmic receptors, which trigger the activation of IRF3 through TLR-independent pathways, and induce type 1 IFN gene transcription (Bowie and Unterholzner, 2008). When type 1 IFNs bind to the IFNAR, downstream signalling can vary depending on cell type, antigen, and IFN subtype. IFNAR dimerization activates JAK1 and TYK2 kinases, which initiate a signalling cascade involving activation of STAT signal transducers. STAT1, STAT2, and IRF9 form the ISGF3 transcriptional activator complex, which binds IFN-stimulated response elements (ISREs) to induce transcription of type 1 IFN-inducible genes (Figure 1-2). An alternative type 1 IFN signalling pathway involves the direct transcription of IFN-inducible genes by STAT4 homodimers. The canonical pathway for type 2 IFN signals through JAK1, JAK2, and STAT1 homodimers, which binds IFN-γ activated site elements (GAS) in the promoter of type 2 IFN-inducible genes. Notably, type 1 IFNs can also activate type 2 IFN-inducible gene transcription, due to shared signalling mechanisms between type 1 and type 2 IFN systems (Theofilopoulos et al., 2005, Ronnblom, 2011).

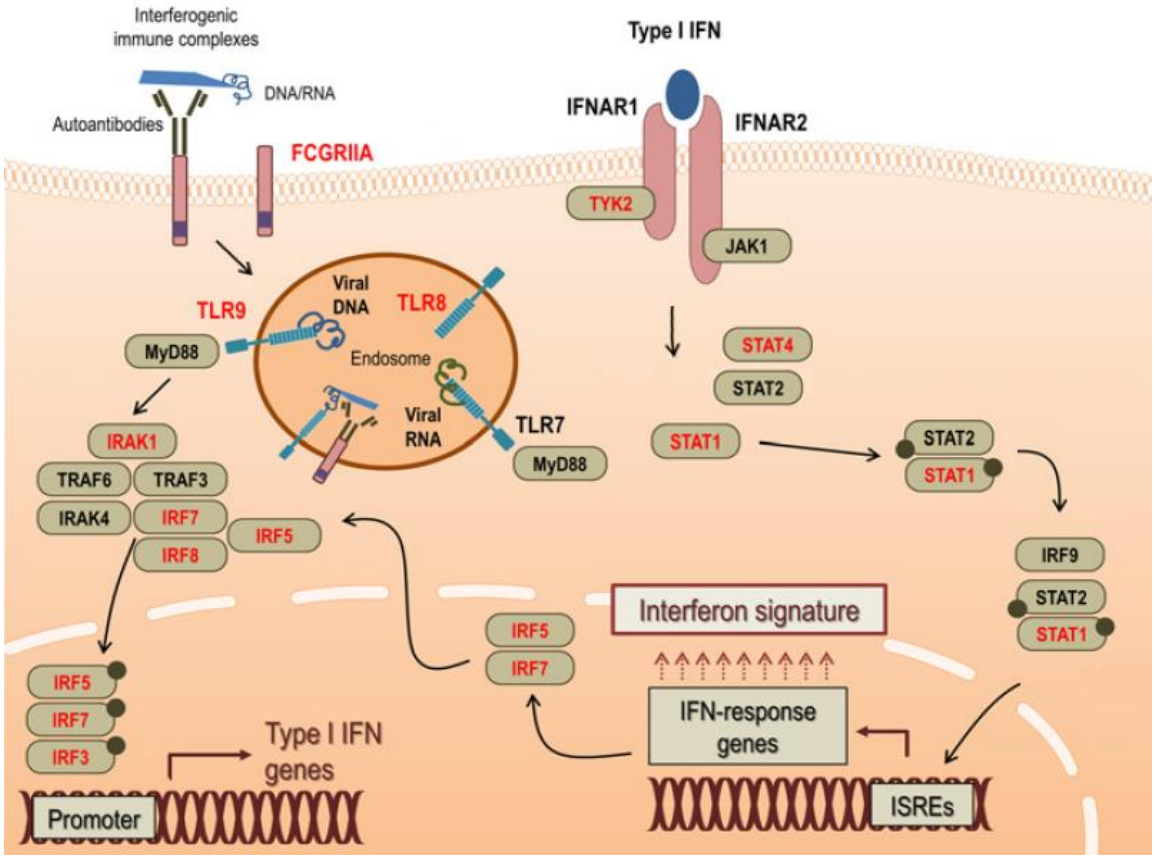


Figure 1-2 TLR-dependent type 1 IFN pathways.

Recognition of nucleic-acid-containing immune complexes by Fcγ receptors (FCGRs) trigger endocytosis and TLR-dependent signalling pathways that induce production of type 1 IFNs (shown on the left). Type 1 IFNs bind to the IFN-α receptor (IFNAR) on the cell surface, and induce signal transduction cascades that activate the expression of several IFN-inducible genes (shown on the right). See text for a detailed description of each pathway. ISRE – IFN-stimulated response element. Figure adapted from Rodriguez-Carrio et al. (2015).

1.3.3 IFN-α in the pathogenesis of SLE

In SLE, the type 1 IFN signature found in patients is mainly associated with IFN-α activity. As illustrated in Figure 1-1, disease pathogenesis is characterized by sustained IFN-α production following antigen exposure, instead of a shutdown of the immune response. The most potent producers of IFN-α are pDCs, which secrete up to 1000 times more IFN-α than other cells when in the presence of viral antigens or ICs. IFN-α is able to breakdown immune tolerance to self by activating APCs (DCs and B cells) and pathways of the adaptive immune system, including up-regulation of HLA class I and costimulatory molecules, cytotoxic CD8+ T cell differentiation, immunoglobulin class switch, and antibody production by plasma cells (Theofilopoulos et al., 2005).

Furthermore, the importance of the type 1 IFN pathway in SLE has been reinforced by genetic studies. The first genes in the IFN-α pathway to show association with SLE were *IRF5* and *TYK2* (Sigurdsson et al., 2005). With the advent of genome-wide association studies (GWAS), further associations with IFN pathway genes have been reported and replicated. These are discussed in more detail in section 1.4.2 of this chapter. Moreover, having high levels of IFN-α is a trait that clusters in families, and is a heritable risk factor for SLE (Niewold et al., 2007). Although it is not known whether increased IFN-α levels are a cause or a consequence of SLE, the association of IFN-α therapy with initiation of autoimmunity, as well as the genetic association between IFN pathway genes and SLE susceptibility, point to a causal role for IFN-α in the disease.

Given the importance of type 1 IFNs in SLE, clinical trials targeting the IFN-α pathway are currently underway, including the use of monoclonal antibodies directed against IFN-α and the IFNAR. A phase I clinical trial using anti-IFN-α antibodies reported a decrease in the type 1 IFN signature in peripheral blood mononuclear cells (PBMCs) and skin, and a reduction in disease activity (Merrill et al., 2011). Given the pleiotropic nature of IFNs, namely their role in anti-viral and anti-tumour defence, one of the potential effects of anti-IFN therapy is an increased risk of viral infections and tumours. Conversely, the functional redundancy between different IFN types and subtypes may result in non-responsiveness to therapy. Targeting molecular components downstream of the IFN receptor may be of more therapeutic benefit; accordingly, further exploration of IFN pathways in SLE is warranted.

1.4 Genetic susceptibility to SLE

SLE has a strong genetic component, as shown by a number of twin and familial aggregation studies. Monozygotic twins show higher concordance rates for the disease (25-70%) compared to dizygotic twins (2-5%, Deapen et al., 1992). Siblings of SLE patients have a higher disease risk when compared to the general population, with a sibling risk ratio (λ) between 8 and 29 (Alarcon-Segovia et al., 2005). One study calculated a heritability of 66% for SLE; however, only 15% of this heritability is explained by genetic factors identified to date (Lawrence et al., 1987, Bentham et al., 2015). In accordance with the common disease-common variant hypothesis,

overall genetic susceptibility to SLE is caused by a combination of multiple common risk variants, each of modest effect size (odds ratio, OR between 1.15 and 2.5). In a minority of SLE cases, genetic susceptibility is attributed to rare, highly penetrant single-gene mutations (monogenic SLE).

The first genetic associations with SLE were identified in 1971 through serological case-control studies. SLE cases were shown to be enriched for the HLA class I alleles (HLA-B8 and HLA-W15) within the MHC region on chromosome 6 (Grumet et al., 1971, Waters et al., 1971). Genome-wide linkage studies subsequently confirmed SLE associations within the MHC region (Gaffney et al., 1998, Gaffney et al., 2000, Forabosco et al., 2006). With the advent of GWAS, variants within the MHC region were shown to confer the greatest genetic risk for SLE (Tsao, 2004, Graham et al., 2008, Bentham et al., 2015). The association of the MHC region with SLE is discussed further in section 1.5 of this chapter, and section 3.1 of Chapter 3.

Over the last decade, GWAS conducted in different populations have provided further insight into the genetic factors conferring susceptibility to SLE, which include several non-HLA loci. GWAS are based on a hypothesis-free approach where the frequencies of common single nucleotide polymorphisms (SNPs) are compared between disease cases and unaffected controls. There are currently more than 40 validated common disease-associated variants, many of which are located within or in proximity to genes functionally involved in SLE-relevant pathways (Bentham et al., 2015). These pathways include clearance of apoptotic cells and ICs, type 1 IFN pathways, and lymphocyte signalling and function. In addition, GWAS have identified a previously unknown role for autophagy and ubiquitination in SLE (Sestak et al., 2007, Flesher et al., 2010, Liu and Davidson, 2012).

1.4.1 Defective apoptosis and clearance of ICs

The complement system is involved in the opsonization and clearance of microbes and apoptotic cells by phagocytosis. One of the strongest genetic associations with SLE is found within *ITGAM*, a gene encoding a subunit of complement receptor 3, CD11b. The R77H variant of *ITGAM* has been shown to impair CD11b-mediated phagocytosis of opsonized targets by monocytes, potentially affecting the clearance of apoptotic cells (Rhodes et al., 2012, Zhou et al., 2013). Furthermore, homozygous loss-of-function (LoF) mutations in several complement

component genes have shown association with monogenic forms of SLE, including *C1Q*, *C1R*, *C1S*, and the MHC region-linked *C2* and *C4* genes (Belot and Cimaz, 2012, Bryan and Wu, 2014). The R77H variant and complement deficiency resulting from LoF mutations have been shown to impair the clearance of apoptotic cells and ICs, thereby exposing self-antigens to immune recognition by autoreactive cells (Belot and Cimaz, 2012, Zhou et al., 2013).

Another group of SLE-risk genes involved in clearance of cellular debris are the FCGRs. FCGRs induce phagocytosis of ICs by binding to the Fc portion of ICs. A copy number variant (CNV) in *FCGR3B* is associated with reduced levels of FCGR3B in neutrophils, which may also impair IC clearance (Yuan et al., 2015). This CNV also generates a *FCGR2C-FCGR2B* hybrid that is expressed ectopically in NK cells, potentially inhibiting NK cell function in SLE (Mueller et al., 2013).

1.4.2 Type 1 IFN pathways

Several SLE susceptibility genes encode components involved in the control of type 1 IFN production and response. Within the pathways leading to IFN production, SLE-associated variants have been identified in the nucleic acid receptor genes *TLR7* and *IFIH1* (MDA5), and in the IFN regulatory factor genes *IRF5*, *IRF7*, and *IRF8*. In the type 1 IFN response pathway, genes encoding the signal transducers *TYK2* and *STAT4* have also shown association with SLE (Mohan and Putterman, 2015, Bentham et al., 2015). Many of the aforementioned SLE-risk variants have been shown to result in gene gain-of-function (GoF). For example, variants in *IFIH1*, *IRF5*, and *STAT4* translate to increased serum IFN-α levels, or increased sensitivity to IFN-α, in either SLE patients or healthy controls (Robinson et al., 2011, Feng et al., 2010, Kariuki et al., 2009).

1.4.3 B cell and T cell signalling

Components of adaptive immunity pathways, particularly those involved in lymphocyte development and function, have also shown association with SLE. Susceptibility variants have been identified in genes pertaining to B cell receptor (BCR) signalling, such as *BLK*, *LYN*, *BANK1*, and *CSK*. The role of co-stimulatory pathways in SLE has been highlighted by the association of variants within *TNFSF4*, a gene encoding a co-stimulatory molecule, OX40L, that acts at the APC-T cell interface. Other variants have been identified in the gene *PTPN22*, which

encodes a protein tyrosine phosphatase that modulates the responsiveness of B cell and T cell receptors; and in transcription factor genes *IKZF1* and *IKZF3*, which play a role in B and T cell development (Gateva et al., 2009, Mohan and Putterman, 2015, Bentham et al., 2015).

1.4.4 Nucleic acid metabolism

Nucleic acid degradation is an important step in programmed cell death (apoptosis), which minimizes exposure of nuclear self-antigens to the immune system. Several genes involved in nucleic acid degradation are associated with monogenic lupus, including the nuclease genes *TREX1*, *DNASE1*, *DNASE1L3*, and the *RNASEH* genes (Belot and Cimaz, 2012). Interestingly, single-gene mutations in components of nucleic acid metabolism are also associated with rare clinical syndromes characterized by high levels of type 1 IFNs. These Mendelian disorders, broadly termed type 1 interferonopathies, exhibit substantial clinical and genetic overlap with SLE, including the presence of a type 1 IFN signature, skin involvement, and inflammation of the central nervous system. Studies of these syndromes have established a correlation between increased serum type 1 IFN levels and LoF mutations in genes encoding nucleases (e.g. *TREX1*, *RNASEH2C*), regulators of genomic stability (e.g. *SAMHD1*), and GoF mutations in nucleic acid receptors (e.g. *IFIH1*, Miner and Diamond, 2014, Crow and Manel, 2015, Kretschmer et al., 2015, Gunther et al., 2015). Thus, these studies have identified the dysregulation of nucleic acid metabolism as an important trigger of IFN-α production in SLE.

1.4.5 Other pathways

Further SLE associations have been reported in other components of innate immunity. These include several genes encoding regulators of NFκB signalling, which occurs downstream of TLR engagement: the *IRAK1* kinase gene, and the ubiquitination regulator genes *UBE2L3*, *TNFAIP3* (A20), and *TNIP1*. Other recently identified SLE-risk loci include the autophagy gene *ATG5*; and the *NCF2* gene, which is thought to be involved in the cytotoxic activity of phagosomes (Gateva et al., 2009, Jacob et al., 2012).

1.4.6 Epigenetics

The epigenome consists of non-coding chromatin modifications, such as methylation and histone modifications, which modulate cell-specific gene expression. A recent methylation study reported hypomethylation of IFN-inducible genes in T cells, B cells, and monocytes from SLE

patients with active and inactive disease. As hypomethylation reflects an open chromatin state, these results indicate an increased sensitivity to IFN-α in SLE patients (Absher et al., 2013). One other epigenetic modification associated with active transcription, histone 4 acetylation, was also shown to be increased in monocytes from SLE patients (Zhang et al., 2010). Changes in microRNA (miRNA) activity have also been found in SLE patients; for example, miRNA-146A, an inhibitor of the type 1 IFN response, is decreased in PBMCs from SLE patients (Tang et al., 2009). Epigenetic studies in SLE have contributed to the identification of relevant disease pathways, and future research in this field is likely to provide a better understanding of the pathogenesis of the disease (Costa-Reis and Sullivan, 2013).

1.4.7 Missing heritability

As most SLE susceptibility variants identified to date confer relatively small effect sizes in overall disease risk, a considerable portion of disease heritability remains undetected. This "missing heritability" may be explained by undiscovered rare variants with large effect sizes, and/or multiple common variants with small effect sizes. Epigenetics, gene-gene interactions (epistasis), and gene-environment interactions involving known and undiscovered SLE-risk variants are also likely to exert strong effects on overall disease risk. Furthermore, sub-phenotype studies may enable the detection of new variants associated with specific clinical features that might be missed in SLE GWAS (Flesher et al., 2010).

1.5 The MHC region

HLA molecules consist of a group of cell surface proteins of the adaptive immune system which present foreign or pathogenic antigens to T cells. These molecules are also involved in maintaining immune tolerance against self by presenting antigens to developing B and T cells. Classical HLA class I molecules are expressed in all nucleated cells, and interact with CD8+ T cells to induce cytotoxic T cell activity. Classical class II molecules are expressed in APCs, namely DCs, B cells, and thymic epithelial cells, and interact with CD4+ T cells to drive antibody-mediated (humoral) immunity. Non-classical HLA class I and II molecules are located intracellularly, and are involved in antigen processing and loading onto classical HLA molecules (Relle and Schwarting, 2012).

HLA molecules are mostly encoded in the MHC region, located on the short arm of chromosome 6 (6p21.3). The MHC region is a gene-dense locus, harbouring 224 genes and pseudogenes. Approximately 40% of genes in the region encode elements involved in immunity. The classical MHC locus spans approximately 3.6 megabases (MB), and is divided in three sub-regions: class I (telomeric), class III, and class II (centromeric). The class I and class II regions contain the HLA genes, which encode classical and non-classical HLA molecules, while the class III region encodes soluble proteins involved in immunity, such as complement factors, cytokines, and heat shock proteins (Figure 1-3, Horton et al., 2004, Relle and Schwarting, 2012).

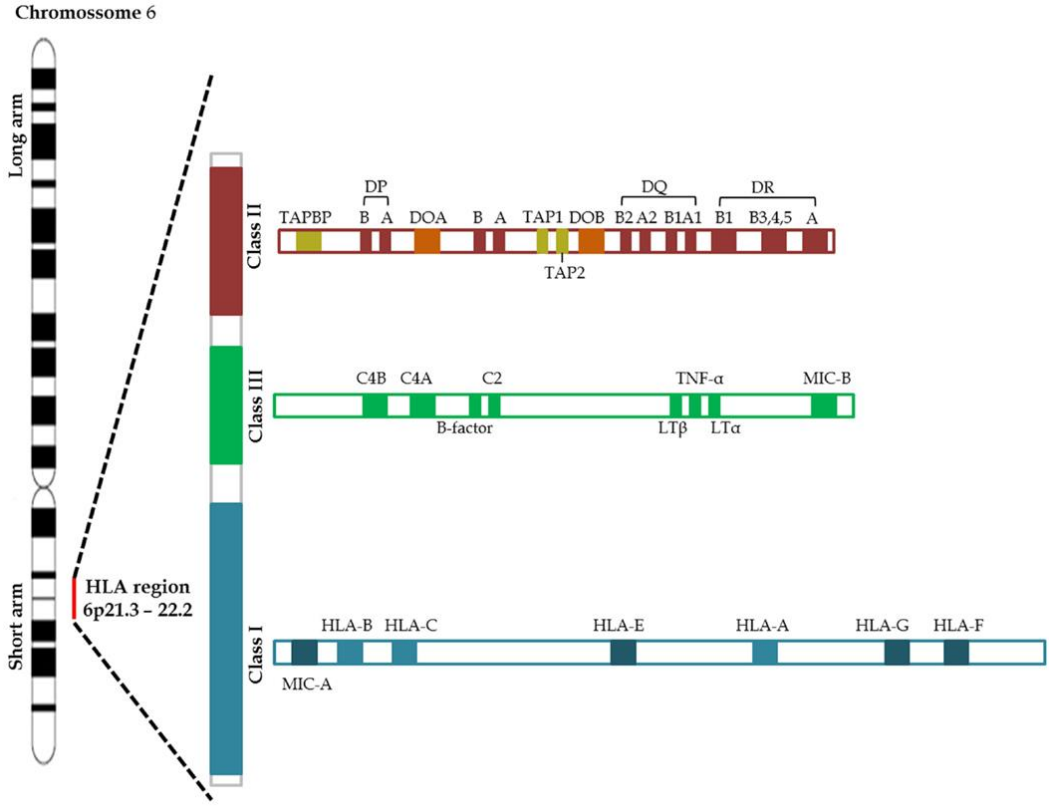


Figure 1-3 Schematic representation of the classical MHC region.

The figure shows the location of classical HLA class I and class II genes, as well a number of relevant class I, II, and III genes in the MHC region. Image adapted from Lima-Junior and Pratt-Riccio (2016).

The MHC region demonstrates a high degree of genetic complexity, owing to extended linkage disequilibrium (LD), the presence of highly polymorphic genes, and copy number variation. This

locus comprises one of the human genome's regions of highest LD, which can extend to immune- and non-immune-related genes outside the classical MHC region. This has led to the characterization of an extended MHC (xMHC) region, which spans 7.6 MB across 421 loci (Horton et al., 2004). The MHC region also harbours the HLA genes, which include the most polymorphic genes in the genome. HLA polymorphism is present in the form of polygeny, allelic diversity, and copy number variation.

Classical HLA class I and II molecules can be encoded by a number of HLA loci. HLA class I molecules are encoded by *HLA-A*, *HLA-B*, and *HLA-C*, while class II molecules are encoded by *HLA-DR*, *HLA-DP*, and *HLA-DQ* genes. Each HLA molecule is a heterodimer composed of two chains: the HLA class I is formed by an α chain associated with β2-microglobulin (β2m), while the HLA class II is formed by an α and a β chain. Whereas α and β chains are encoded by classical HLA genes, the β2m gene is located outside the MHC region (Figure 1-4). In HLA molecules, the specificity of the peptide binding groove is determined by hypervariable regions within the α and β chains. In HLA class I, the hypervariable regions are encoded within exons 2 and 3 of HLA class I genes; in HLA class II, these regions are encoded within the exon 2 of HLA class II genes (except *HLA-DRA*, which is not polymorphic). The wide allelic diversity of the classical HLA genes means that most individuals in the population are heterozygous at each locus. Therefore, one individual can have up to six isotypes of HLA class I and class II molecules (2 alleles × 3 loci), thus enabling protection against a wide range of pathogens (Lahita, 2004, Relle and Schwarting, 2012).

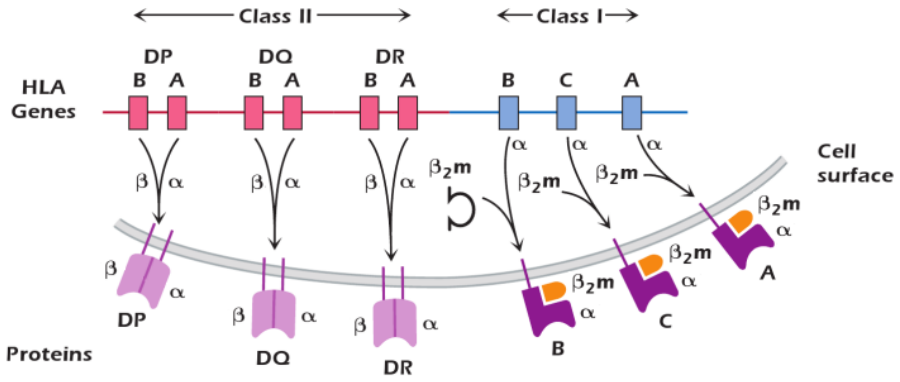


Figure 1-4 Illustration of the classical human HLA genes and molecules.

The genes for classical HLA class I (A, B, and C) and class II (DP, DQ, and DR) molecules are encoded within the MHC region, represented at the top in centromeric to telomeric order. β2-microglobulin (β2m) is encoded outside the MHC region. In HLA class I, the α chain is hypervariable, whereas β2m is not polymorphic. In HLA class II, the α and β chains of HLA-DP and HLA-DQ, and the β chain of HLA-DR are hypervariable, whereas the α chain of HLA-DR is not polymorphic. Figure adapted from Coico and Sunshine (2015).

The class II *HLA-DRB* locus is one of two regions of the MHC that exhibit gene copy number variation. *DRB1* haplotypes can carry one or two functional *DRB* genes (*DRB1* and *DRB3*, *DRB4*, *DRB5*, or none), and one to three pseudogenes (*DRB9* and *DRB2*, *DRB6*, *DRB7*, *DRB8*, or none), as illustrated in Figure 1-5 (Lahita, 2004). The polymorphic nature of HLA genes is the result of evolutionary pressure from rapidly mutating pathogens with potential to evade the immune system. The current models explaining allelic diversity at the MHC postulate that heterozygote individuals have the ability to recognize more pathogens than MHC-homozygous individuals (heterozygote advantage), and that the formation of new rare allelic combinations provides an evolutionary advantage against pathogens at the population-level (rare allele advantage, Relle and Schwarting, 2012).

Figure 1-5 Haplotypes for the class II *HLA-DRB* locus.

The figure shows the structure of five *DRB* haplotypes, which include expressed genes (in green) and pseudogenes (in blue). Image exported from Lahita (2004).

1.5.1 Disease associations with the MHC region

The MHC region is one of the most extensively studied loci in the human genome, due to its strong genetic association with disease. Hundreds of diseases and traits have shown association with variants within the region, including multiple autoimmune diseases, inflammatory diseases, infectious diseases, transplant rejection, and adverse drug reactions (Horton et al., 2008). As previously mentioned, the MHC region confers the strongest risk for SLE in virtually every population studied to date, and is currently thought to explain approximately 4% of disease heritability (Dr David Morris, King's College London; personal communication). However, the causal variants underlying disease association at the MHC are difficult to dissect, owing to tight LD and high polymorphism in the region, copy number variation, epistatic interactions between HLA and/or non-HLA alleles, and the clinical heterogeneity of SLE probands (Lincoln et al., 2009, Morris et al., 2012). Low recombination frequencies at the MHC are attributed to past selective pressures which favoured the preservation of specific allelic combinations that provided protection from infectious diseases (Gluckman and Hanson, 2004, Relle and Schwarting, 2012). This has led to the conservation of extended ancestral haplotypes which can span up to 4 MB, encompassing part of class I, the whole of class III, and part of the class II region (from *HLA-B* to *HLA-DQ*, Horton et al., 2008, Bishof et al., 1993). Ancestral MHC haplotypes are common in the population (frequency ≥ 0.5%), and together they make up about half of all MHC haplotypes in Europeans. Importantly,

most disease associations with the MHC region involve markers of extended ancestral haplotypes (Larsen et al., 2014).

The current human reference sequence for the MHC region is derived from a lymphoblastoid cell line (LCL) homozygous for the ancestral *HLA-B*07-DRB1*15:01* haplotype (PGF). In order to characterize haplotype-specific variation, the MHC Haplotype Project annotated the sequences of seven additional MHC-homozygous LCLs (Table 1-1, Horton et al., 2008). The European ancestral MHC haplotypes harboured by these cell lines were selected for sequencing owing to their association with disease risk or protection, including SLE, type 1 diabetes, and multiple sclerosis.

Table 1-1 HLA haplotypes and alleles of eight MHC-homozygous cell lines sequenced by the MHC Haplotype Project (Horton et al., 2008).

Cell Line	Haplotype	HLA-A	HLA-B	HLA-C	HLA-DRB1	HLA-DQB1	HLA-DPB1
PGF	A3-B7-DR15	03:01	07:02	07:02	15:01	06:02	04:01
COX	A1-B8-DR3	01:01	08:01	07:01	03:01	02:01	03:01
QBL	A26-B18-DR3	26:01	18:01	05:01	03:01	02:01	02:02
APD	A1-B60-DR13	01:01	40:01	06:02	13:01	06:03	04:02
DBB	A2-B57-DR7	02:01	57:01	06:02	07:01	03:03	04:01
MANN	A29-B44-DR7	29:02	44:03	16:01	07:01	02:02	02:01
SSTO	A32-B44-DR4	32:01	44:02	05:01	04:03	03:05	04:01
MCF	A2-B62-DR4	02:01	15:01	03:04	04:01	03:01	04:02

The most significant associations with SLE in Europeans consistently arise from the ancestral *B*08-DRB1*03:01* haplotype, which is tagged by the *DRB1*03:01* allele (Grumet et al., 1971, Goldberg et al., 1976). The second most consistently replicated association with SLE in Europeans arises from the ancestral haplotype tagged by the *DRB1*15:01* allele (*B*07-DRB1*15:01*, Uribe et al., 2004, International MHC and Autoimmunity Genetics Network et al.,

2009). SLE associations with other *DRB1* alleles, including *DRB1*04*, *DRB1*08*, *DRB1*09:01*, *DRB1*13*, and *DRB1*14*, have been reported in different populations (Kapitany et al., 2009, Uribe et al., 2004, Shimane et al., 2013, Furukawa et al., 2014, Kim et al., 2014). The association of classical HLA genes with SLE, as well as other autoimmune diseases, is likely to be explained by the role of HLA molecules in antigen presentation. Disease-associated alleles may translate to HLA amino acid variations that result in poor presentation of self-antigens to T cells during central tolerance, thereby disturbing negative selection thresholds for autoreactive T cells in the thymus (Wucherpfennig and Sethi, 2011).

In recent years, targeted studies have attempted to fine-map MHC associations in SLE, in order to identify the causal variants. Implementation of high-density SNP genotyping, trans-ancestral mapping, and meta-analysis have enabled the identification of independent SLE associations at the MHC, including risk and protective variants in *HLA-G* (class I), *HLA-DPB1* (class II), *SKIV2L* (class III), and *MSH5* (class III, Fernando et al., 2007, Ruiz-Narvaez et al., 2011, Fernando et al., 2012). Interestingly, after conditioning on *DRB1* signals, these studies identified associations with the classical class II *DPB1* gene in Spanish (rs3117213, risk, OR = 1.76), Filipino (rs2071351, protective, OR = 0.48), and African-American (rs2071349, risk, OR = 1.53) SLE cohorts. Similarly, an SLE-protective variant at *MSH5*, rs409558, was associated in both Spanish and Filipino populations independently of *DRB1* signals (OR ~ 0.56). Since the aforementioned variants in *DPB1* and *MSH5* are non-coding, the causal mechanisms underlying these associations are likely to be independent from HLA amino acid variation. However, the independence of these signals from HLA amino acids has not been yet established.

1.5.2 *DRB1*03:01* haplotypes

The *DRB1*03:01* allele is common in European populations (frequency of 10%). It tags two major haplotypes: one common in Northern Europeans (*B*08-DRB1*03:01*, present in COX cell lines), and the other common in Southern Europeans (*B*18-DRB1*03:01*, present in QBL cell lines, Table 1-1). *DRB1*03:01* haplotypes are some of the longest haplotypes found in the human genome, and can span up to 4 MB of the MHC region, including the majority of class I and II, and all of class III. Sequencing and annotation of the *B*08-DRB1*03:01* haplotype,

carried out by the MHC Haplotype Project revealed that this haplotype spans 311 loci and 1,191 variants (Horton et al., 2008).

In addition to SLE, *DRB1*03:01* haplotypes have been associated with multiple autoimmune as well as non-autoimmune diseases. Among the autoimmune diseases associated with this haplotype are primary and secondary Sjögren's syndrome, rheumatoid arthritis, idiopathic inflammatory myopathies, and type 1 diabetes. A full list of diseases associated with *DRB1*03:01* haplotypes is presented in Table 1-2.

Table 1-2 Diseases that have shown association with *HLA-DRB1*03:01* haplotypes.

Disease	**Reference**
Autoimmune	
Systemic lupus erythematosus	Graham et al. (2002), Morris et al. (2012)
Primary Sjögren's syndrome	Gottenberg et al. (2003)
Rheumatoid arthritis	Vignal et al. (2009)
Multiple sclerosis	de la Concha et al. (2012)
Type 1 diabetes	Noble and Valdes (2011)
Idiopathic inflammatory myopathies	O'Hanlon et al. (2006), Miller et al. (2015)
Graves' disease	Maciel et al. (2001)
Autoimmune hepatitis	Oliveira et al. (2011)
Celiac disease	Jores et al. (2007)
Dermatitis herpetiformis	Wilson et al. (1995)
Systemic sclerosis	Kallenberg et al. (1981)
Sarcoidosis	Grubic et al. (2007)
Myasthenia gravis	Degli-Esposti et al. (1992)
Lambert-Eaton myasthenic syndrome	Parsons et al. (2000)
Immunoglobulin A deficiency	Volanakis et al. (1992)
Non-autoimmune	
Ovarian cancer	Kubler et al. (2006)
Human immunodeficiency virus infection	Steel et al. (1988)
Common variable immune deficiency	Volanakis et al. (1992)
Non-Hodgkin lymphoma	Abdou et al. (2010)

Several studies have demonstrated strong association between *DRB1*03:01* haplotypes and the autoantibody subsets anti-Ro and anti-La in European SLE (OR = 3.40-7.50, Galeazzi et al., 2002, Christian et al., 2007, Morris et al., 2014). This suggests that *DRB1*03:01* haplotypes strongly increase the risk of specific sub-phenotypes comprising anti-Ro and anti-La antibodies, rather than predisposing uniformly to SLE. Given that HLA class II presentation to T cells is required for the initiation of antibody-mediated immunity, these findings strongly support a direct role for the *DRB1*03:01* allele in the pathogenesis of the disease.

Several non-HLA genetic variants within *DRB1*03:01* haplotypes have been associated with SLE, suggesting that multiple risk factors within these haplotypes may contribute to increased susceptibility to SLE, in addition to the risk conferred by HLA alleles. Examples of such *DRB1*03:01*-linked SLE associations that may be functionally relevant include the class III *TNF* -308/G>A promoter variant, the class III *TNFB*1* variant, and the *C4A* and *C4B* null alleles (*C4A*Q0* and *C4B*Q0*, Wilson et al., 1993, Bettinotti et al., 1993, Boteva et al., 2012). *TNF* and *TNFB* (or *LTA*) both encode cytokines from the tumour necrosis factor (TNF) family (TNF-α and TNF-β, respectively). TNF has been shown to be both beneficial and aggravating in murine models of lupus (Aringer and Smolen, 2008). Similarly, clinical trials testing the efficacy of TNF-α inhibition have shown contradictory outcomes in lupus; accordingly, the role of TNF in disease is still unclear (Takahashi et al., 2008, Aringer et al., 2009, Uppal et al., 2009). Interestingly, the SLE-risk -308G>A promoter variant in *TNF* correlates with increased expression levels of *TNF* in blood leukocytes and other cell types (Wilson et al., 1993, Helmig et al., 2011, Abraham and Kroeger, 1999). Complement component C4 is encoded by the copy number variable genes *C4A* and *C4B*. Ancestral *B*08-DRB1*03:01* and *B*18-DRB1*03:01* haplotypes harbour only one *C4* copy (*C4A* or *C4B*, respectively). Even though partial genetic *C4* deficiency does not independently predispose to SLE, complete homozygous *C4* deficiency, and complete C4 protein deficiency, are known risk factors for lupus (Boteva et al., 2012, Wu et al., 2009b, Yang et al., 2004). Thus, partial genetic *C4* deficiency may increase disease risk when combined with other risk factors present on *DRB1*03:01* or non-*DRB1*03:01* haplotypes.

1.6 Genetics of gene expression in immunity

GWAS in human complex diseases have identified hundreds of associated variants, of which approximately 90% are non-coding (Hindorff et al., 2009). One of the possible functional consequences of these variants is regulation of gene expression, for example, by altering messenger RNA (mRNA) expression levels, alternative splicing, or miRNA activity. The level of expression of a transcript in cells or tissues can be considered a genetically-determined quantitative trait, similarly to traditional quantitative phenotypes such as weight and height. Genetic variants that alter gene expression levels are known as expression quantitative trait loci (eQTLs), and can be mapped by testing the association between transcript levels and specific genetic markers in a cohort of individuals. eQTLs can be analysed for a single locus, or by generating genome-wide expression profiles using microarray or RNA sequencing technologies. Commonly, a genome-wide variation profile is also generated for the cohort, using high-density SNP genotyping platforms.

Regulatory variants are defined as *cis*- or *trans*-acting, depending on their genomic distance from the target gene. *Cis*-acting variants (or *cis*-eQTLs) are variants that alter the expression of nearby genes, and are conventionally defined as being located at a distance of up to 1 MB from the target gene. *Trans*-eQTLs are defined as variants that affect distant genes, or genes on a different chromosome. Genes underlying *trans*-eQTLs are assumed to encode *trans*-acting regulatory factors that bind *cis*-elements of other genes. Thus, *trans*-eQTLs may represent the location of transcription regulators (such as transcription factors) that alter the expression of one or more functionally-related genes, and have the potential to expose new biologically-relevant pathways. In the last decade, several eQTL studies have successfully identified multiple *cis*-associations (Westra and Franke, 2014, Fairfax and Knight, 2014). In contrast, *trans*-eQTLs have been more difficult to detect, due to their complexity and the need to correct for large numbers of statistical tests.

1.6.1 Context-dependency of eQTLs

Early eQTL studies were often conducted in heterogeneous cell populations such as whole blood, PBMCs, or skin biopsies. These studies helped to elucidate the global eQTL landscape in these cell populations and tissues, as they translated *in vivo* cell states and cell-to-cell

interactions. However, the importance of biological context for the detection of eQTLs, including cell type, stimuli, time point, and cell environment, has recently been highlighted (Fairfax and Knight, 2014). eQTLs from less abundant cell types (such as B cells in whole blood or PBMC samples) may therefore not be detected in data sets generated from heterogeneous cell populations, due to signal saturation from the most common cell types. eQTL studies designed to explore cell type-specific eQTLs have been growing in number, and have been conducted using purified immune cell subsets, such as *ex vivo* monocytes, B cells, CD4+ T cells, CD8+ T cells, and neutrophils. eQTLs may be present in one cell type but not in others, or may have opposing effects in different cell types (Fairfax et al., 2012, Westra et al., 2015, Naranbhai et al., 2015).

Additionally, eQTLs may only be detected when cells are exposed to specific stimuli – such eQTLs are named response eQTLs (reQTLs). For example, one study conducted using *ex vivo* B cells found different eQTLs compared to studies using LCLs, which are derived from Epstein-Barr virus-transformed B cells (Fairfax et al., 2012). Other studies have reported several novel reQTLs following stimulation with lipopolysaccharide (LPS), IFN-β, IFN-γ, and infection with the influenza virus in monocytes and/or DCs (Fairfax et al., 2014, Lee et al., 2014). It is possible that immune stimulation of cells may facilitate the discovery of eQTLs by homogenising gene expression. Furthermore, reQTL studies may also enable the discovery of *trans*-eQTLs, which tend to be more sensitive to context specificity than *cis*-eQTLs. The mechanisms underlying eQTLs and their regulation of gene expression may involve the formation/alteration of transcription factor or miRNA binding sites (Cham et al., 2012, Luo et al., 2011), non-synonymous variants that result in a shortened transcript life-span or nonsense-mediated RNA decay, or splice site variants that alter the levels of spliced transcript isoforms (Lalonde et al., 2011).

1.6.2 Methods for quantifying gene expression

The development of high-throughput genotyping and expression platforms has enabled the generation of high-dimensional SNP genotype and gene expression data, which can be used for genome-wide eQTL analyses. Several technologies can be employed to measure transcript levels, including quantitative reverse-transcription polymerase chain reaction (qPCR, for single

locus analyses), or gene expression microarrays and next generation sequencing (for generation of genome-wide expression profiles).

Commonly used gene expression platforms include the Affymetrix GeneChip array and the Illumina BeadChip array (Turnbull et al., 2012). Traditional gene expression microarrays target the 3' end of annotated genes to measure overall expression levels, but provide little information on alternatively spliced transcripts. Affymetrix platforms target each gene using multiple 25-base long oligodeoxynucleotide probes, in contrast to other platforms, where each gene is generally targeted by one probe. In principle, Affymetrix whole-transcriptome arrays provide the most reliable microarray technology, due to their dense genomic coverage. The Affymetrix Human Exon 1.0 ST array (or exon array) contains probe sets that specifically target every putative exon, which provides a more accurate quantification of gene expression.

Seventy-five per cent of eukaryotic genes and 95% of multi-exon genes are estimated to undergo alternative splicing, which enables the generation of different mRNAs and, subsequently, protein diversity from a single gene (Pan et al., 2008, Tian et al., 2011). Exon arrays are one of the methods which offer insights into alternative splicing events and alternative transcription start sites (Lockstone, 2011). Alternatively, cell type-specific splicing patterns can be defined using splice junction arrays, such as the Affymetrix GeneChip Human Transcriptome Array 2.0, which targets annotated exon-exon junctions specific to individual transcript splice isoforms. RNA sequencing platforms, such as the Roche 454, Illumina HiSeq, and Life Technologies SOLiD platforms, are also increasingly being used to quantify gene expression (Wall et al., 2009). These technologies have the advantage of measuring all annotated and non-annotated transcripts in a given sample, thus providing unbiased information. However, the generation of gene expression profiles using different platforms raises a caveat regarding the reproducibility of gene expression data between different studies.

1.6.3 Functional applications of eQTL studies

As GWAS do not provide a functional characterization of disease-associated loci, gene expression levels can be used as intermediate phenotypes to investigate the genetic basis of complex disease. Several eQTL studies conducted in *ex vivo* myeloid and lymphoid cells have investigated eQTLs in GWAS variants associated with autoimmune and inflammatory disorders,

including SLE. A study by Fairfax and colleagues found that autoimmune disease-associated SNPs are enriched for eQTLs genome-wide, suggesting a functional role for these variants in disease risk (Fairfax et al., 2012). Furthermore, SLE-associated variants were found to be enriched for B cell-specific eQTLs compared to monocytes, which is consistent with the importance of B cells in the pathogenesis of the disease. Various *cis*-eQTLs have been reported for SLE-associated variants genome-wide, including those within *IRF5*, *IRF7*, *BLK*, and *UBE2L3* (Fairfax et al., 2012, Bentham et al., 2015). For example, an SLE-risk haplotype at *IRF5* has been shown to correlate with increased levels of *IRF5*, which is potentially caused by an inserted transcription factor binding site for SP1 (Cham et al., 2012). This exemplifies how eQTL studies may help to elucidate the functional mechanisms underlying GWAS associations with complex disease.

Since disease-causing variants might only be active in the presence of a specific cell stimulus, a number of studies have investigated reQTLs in disease-relevant contexts. One reQTL involving the *IRF7* variant rs12805435, which is in moderate LD with the SLE-risk SNP rs4963128, was described by Lee and colleagues following stimulation of DCs with the influenza virus (Lee et al., 2014). After exposure, rs12805435 correlated in *cis* with *IRF7* expression, and in *trans* with expression levels of IFN-α genes, which is consistent with the association of viral infections and type 1 IFNs with the development of SLE in susceptible individuals.

1.7 Summary and project aims

SLE is an autoimmune disease characterized by up-regulation of type 1 IFNs and B cell hyperactivity. Genetic association studies support the dysregulation of the IFN pathway and antigen presentation as the main drivers of disease pathology. The MHC region confers the greatest genetic risk for SLE, but fine-mapping causal variants has been challenging due to the presence of highly polymorphic genes and strong LD. Multiple independent associations with SLE risk and protection have been identified at the MHC, including the *DRB1*03:01* and *DRB1*15:01* ancestral haplotypes, *DPB1* variants in class II, and *MSH5* variants in class III. However, the most statistically significant associations with SLE in Europeans consistently arise within extended *DRB1*03:01* haplotypes, which encompass thousands of genetic variants.

Several causal variants within *DRB1*03:01* haplotypes may therefore contribute to disease risk, possibly through independent mechanisms. Furthermore, several autoimmune diseases have shown association with both *DRB1*03:01* haplotypes and up-regulation of type 1 IFNs; hence, one can hypothesise a gene-environment interaction between these SLE susceptibility factors.

One of the causal mechanisms underlying the aforementioned genetic associations with SLE may be the regulation of gene expression in specific biological contexts. B cells represent an ideal cell type in which to study SLE-associated eQTLs, due to their central role in the pathogenesis of SLE. This project aims to:

i) investigate whether the MHC variants underlying the *DRB1*03:01*, *DRB1*15:01*, *MSH5*, and *DPB1* SLE associations alter gene expression levels in resting LCLs and *ex vivo* B cells, using publicly-available data sets generated from healthy individuals;

ii) explore the effects of IFN-α on gene expression and alternative splicing in *ex vivo* B cells from healthy female twins, using data generated from Affymetrix Human Exon 1.0 ST arrays;

iii) validate the eQTLs identified in aim (i) for the *DRB1*03:01*, *DRB1*15:01*, *MSH5*, and *DPB1* haplotypes in resting *ex vivo* B cells, and investigate whether these haplotypes harbour reQTLs following stimulation with IFN-α;

iv) explore the effect of *DRB1*03:01* haplotypes on alternative splicing in *ex vivo* B cells at rest, and following stimulation with IFN-α;

v) investigate whether there are haplotype-environment interactions between *DRB1*03:01* haplotypes and IFN-α in *ex vivo* B cells at the mRNA level.

Chapter 2 - Materials and methods

2.1 Gene expression datasets

Expression quantitative trait loci (eQTLs) within haplotypes associated with systemic lupus erythematosus (SLE) were investigated in three publicly-available gene expression data sets.

2.1.1 LCL dataset from the International HapMap Project

Gene expression data generated from lymphoblastoid cell lines (LCLs) in the study by Stranger and colleagues were obtained from the Gene Expression Omnibus repository (accession number GSE6536, Stranger et al., 2007). The samples were derived from 60 healthy individuals of Northern and Western European origin from Utah, U.S.A. (CEU), recruited by the International HapMap Project. Expression data were generated using the Illumina Sentrix Human-6 Expression BeadChip platform.

2.1.2 LCL dataset from the MuTHER study

Gene expression data from LCLs generated for the Multiple Tissue Human Expression Resource (MuTHER) study were accessed using the Genevar resource (Grundberg et al., 2012, http://www.sanger.ac.uk/resources/software/genevar). The samples were derived from 777 healthy female individuals from the United Kingdom (U.K.), recruited from the TwinsUK adult registry (Spector and Williams, 2006). Expression data were processed using the Illumina Human HT-12 v3 BeadChip platform.

2.1.3 *Ex vivo* B cell dataset

Gene expression data generated from *ex vivo* B cells in the study published by Fairfax and colleagues were obtained from ArrayExpress (accession number E-MTAB-945, Fairfax et al., 2012). Cells were purified by positive selection from peripheral blood mononuclear cells (PBMCs) of 281 healthy individuals from Oxford, U.K. The data were processed using the Illumina Human HT-12 v4 BeadChip platform.

2.2 Samples for generation of gene expression data

2.2.1 Cohort from the TwinsUK adult registry

Fifty healthy female participants were recruited from the TwinsUK adult resource at St. Thomas' Hospital (Spector and Williams, 2006). Genotype data were provided by the TwinsUK registry for 7000 individuals. Individuals were selected based on the presence of the *DRB1*03:01* allele (using the tag single nucleotide polymorphism (SNP) rs2187668), and the *MSH5* protective haplotype (using the tag SNP rs409558). Thus, the rs2187668 SNP (minor allele frequency of 10%) would have approximately 70 homozygotes available for recall, and the rs409558 SNP (minor allele frequency of 15%) would have approximately 160 homozygotes available for recall. Whole blood samples from twins were collected in two different batches, one year apart. Written consent was obtained from all study participants prior to sample collection.

2.2.2 MHC-homozygous LCLs from the MHC Haplotype Project

Seven LCLs from the MHC Haplotype Project (Horton et al., 2008) were obtained from the European Collection of Authenticated Cell Cultures (https://www.phe-culturecollections.org.uk/collections/ecacc.aspx). All seven LCLs are homozygous at the major histocompatibility complex (MHC) region. Human leukocyte antigen (HLA) genotypes for each cell line are shown in Chapter 1, Table 1-1. The LCLs studied were COX, QBL, PGF, DBB, APD, SSTO, and MANN.

2.3 Generation of gene expression data

Gene expression microarray data were generated from MHC-homozygous LCLs, and from twins' *ex vivo* B cells, both at rest and following stimulation with interferon (IFN)-α, as illustrated in Figure 2-1. The processing of samples prior to array hybridization for 15 of 50 individual twins, and from all seven LCLs, was conducted by Ms Lora Boteva under the supervision of Dr Michelle Fernando (Department of Medical and Molecular Genetics, King's College London). For LCL samples, PGF, COX, QBL, and DBB duplicates were cultured and processed in a different batch to the other samples. Twins' *ex vivo* B cell samples were processed in two batches, and samples from one monozygotic twin pair were processed in each batch.

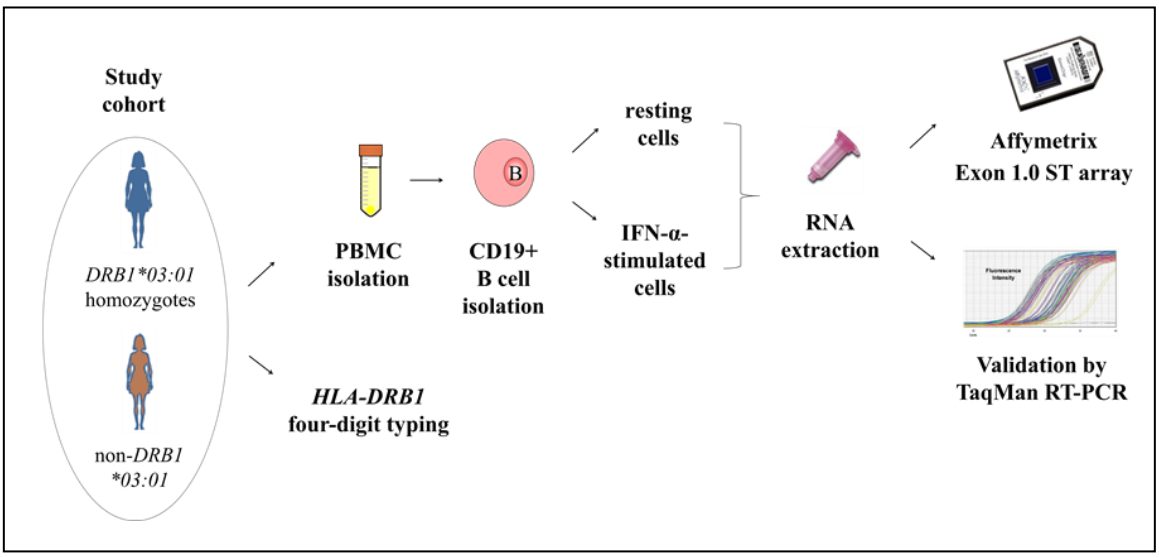


Figure 2-1 Workflow demonstrating the steps taken to generate gene expression data from twins' *ex vivo* B cells obtained from whole blood.

2.3.1 PBMC isolation

PBMCs were isolated from 60 ml whole blood diluted 1:1 in phosphate-buffered saline (PBS, Sigma-Aldrich). 25 ml aliquots were layered onto 15 ml Histopaque-1077 Hybri-Max (Sigma-Aldrich) and centrifuged with no brake at 800 × g for 20 minutes (min) at room temperature (RT). The layer containing PBMCs (buffy coat) was isolated and washed twice with 40 ml PBS/0.5% bovine serum albumin (BSA, Sigma-Aldrich) using centrifugation at 300 × g for 10 min at RT. A final wash and centrifugation at 100 × g for 10 min, RT, was carried out to remove excess platelets. Around 40 - 60 million PBMCs were obtained per blood sample.

2.3.2 DNA extraction

Genomic deoxyribonucleic acid (DNA) was extracted from the cell pellets which remained following PBMC extraction from the 50 twins. Red blood cells were lysed by incubation with ice cold water for 5 min, followed by centrifugation at 300 × g for 10 min, 4 °C. The incubation and centrifugation steps were repeated once. White blood cells were lysed by incubation with hypotonic lysis buffer (1% NP-40, 50 mM Tris-Cl pH 7.5, 150 mM NaCl, Fisher Scientific) for 5 min on ice, followed by centrifugation for 10 min, 4 °C. Pellets were washed twice with PBS, and incubated overnight with genomic lysis buffer (10mM Tris pH 8.0, 10mM NaCl, 10mM EDTA pH8.0, 1% SDS, Fisher Scientific) and Proteinase K (Qiagen). The DNA was purified using

phenol (Fisher Scientific) and chloroform (Fisher Scientific), and was ethanol-precipitated (Fisher Scientific). DNA purity was assessed using the Nanodrop 3300 fluorospectrometer (Thermo Scientific).

2.3.3 *HLA-DRB1* typing

Forty-six of fifty DNA samples from twins participating in the study were successfully typed for *HLA-DRB1* by Dr Robert Collins (Clinical Transplantation Laboratory, Guy's Hospital). Four-digit typing was performed using the SeCore DRB1 Locus Exon 2 & Exon 3 Sequencing Kit (Life Technologies). Using this kit, accurate *DRB1* typing is achieved through amplification and bidirectional sequencing of *DRB1* exon 2 and exon 3. The location of primers within the introns flanking exons 2 and 3 enables locus-specific amplification, and targeting of all *DRB1* alleles.

2.3.4 B cell isolation

PBMCs extracted from whole blood were washed with 40 ml of B cell isolation buffer (PBS, 2% FBS, 2 mM EDTA), and centrifuged at 300 × g for 10 min, RT. CD19+ B cells were isolated by negative selection using the Dynabeads Untouched Human B Cells Kit (Invitrogen). This kit achieves negative selection of CD19+ B cells by depleting non-B cells from samples. Samples were incubated with an antibody mix designed to bind all blood cells except B cells. Antibody-binding Dynabeads were added to samples for a short period, to allow binding to antibody-labelled cells. Bead-bound cells were then removed using a magnet and discarded, leaving untouched CD19+ B cells in the samples.

2.3.5 Cell culture and IFN-α stimulation

Ex vivo B cells and LCLs were cultured in RPMI 1640 medium (GIBCO), 20% foetal bovine serum (FBS, Life Technologies), L-glutamine (200 µM, GIBCO) and 1% penicillin/streptomycin (GIBCO), in a cell culture incubator at 37°C and 5% CO_2. LCLs were thawed from liquid nitrogen and cultured at a density of 5×10^{-5} cells/ml in T75 flasks (Nunco). After reaching exponential growth, 10 million LCLs at rest and 1.5 - 3 million *ex vivo* B cells were harvested. A similar number of cells were stimulated with IFN-α 2b (1000 units/ml, Bio-Techne) for 6 hours (h) in 10 ml of culture medium before harvesting.

2.3.6 RNA extraction

Ribonucleic acid (RNA) was isolated from 1.5 - 3 million *ex vivo* B cells or LCLs, using the RNeasy Mini kit (Qiagen), according to the manufacturer's instructions. This RNA isolation procedure involves a cell lysis step, and on-column RNA-purification. Samples were first lysed, homogenized and mixed with ethanol to provide ideal binding conditions. The lysate was then loaded onto a RNeasy silica membrane, which binds and retains RNA, while allowing all other lysate components to wash through. To remove residual DNA contamination, an on-column DNase I treatment step was included, using the RNase-Free DNase Set (Qiagen). 80 µl of the DNase I incubation mix was added to the membrane. The RNA was then eluted from the membranes in RNase-free water. RNA integrity was assessed in a Nanodrop 3300 fluorospectrometer (Thermo Scientific) and Agilent 2100 Bioanalyzer (Agilent), using the RNA 6000 Pico Kit (Agilent). Samples with a RNA Integrity Number (RIN) > 8 were taken forward for analysis.

2.3.7 cDNA synthesis and array hybridization

Complementary DNA (cDNA) was synthesised from 50 ng of RNA using the High Capacity RNA-to-cDNA Kit (Applied Biosystems). cDNA from each sample was labelled with a fluorescent dye and hybridised to Affymetrix Human Exon 1.0 ST arrays. Data were processed using the GeneChip Scanner 3000 7G (Affymetrix), according to the manufacturer's instructions. This work was performed by Mr Erick Nasser, Dr Fei Wong, and Dr Estibaliz Aldecoa-Otalora Astarloa, under the supervision of Dr Matt Arno (Genomics Centre, King's College London).

2.3.8 The Affymetrix Human Exon 1.0 ST array

The Affymetrix Human Exon 1.0 ST array (henceforth referred to as exon array) is a platform suited for the analysis of gene expression and alternative splicing. The array contains over 5.5 million probes, grouped into 1.4 million probe sets. Each exon is targeted by one probe set, which usually comprises a set of four probes. Longer exons or extended untranslated (UTR) regions may have more than one probe set. Probes are designed to target the reference human MHC sequence, obtained from the PGF cell line. Following hybridization of cDNA to the array, expression data is obtained by fluorescence-based detection. Signal intensities are captured by camera, quantified, and stored as .CEL files. The summarization of raw intensity data can

provide measures of two different kinds of information: i) overall gene expression, summarized as transcript cluster data (gene-level), and ii) exon-specific expression, summarized as probe set data (exon-level). Data quality control (QC) and final summarization were carried out using the probe set and transcript cluster annotation release 33.1 (annotation files dated 12/10/12, GRCh37 build).

2.4 Quality control of exon array expression data

The QC steps described in this section were applied to the raw intensity signal data from exon arrays, and are summarised in Figure 2-2.

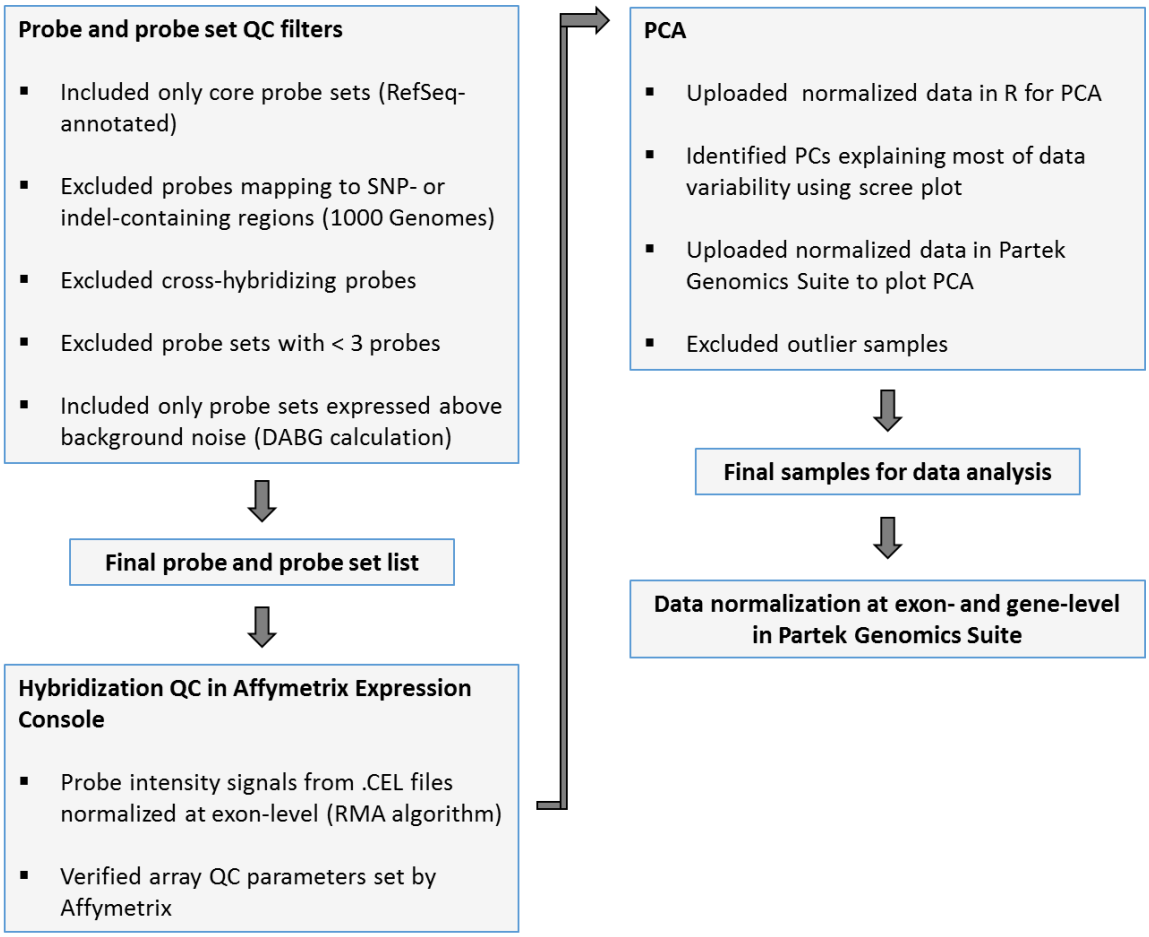


Figure 2-2 Pipeline of QC steps applied to intensity signal data obtained from Affymetrix Human Exon 1.0 ST array.

2.4.1 Probe and probe set quality control filters

Probe and probe set filters were applied to the data, as recommended by Lockstone (Lockstone, 2011). All probe sets targeting RefSeq-annotated RNA transcripts ("core" probe sets) were included. Probes mapping to regions containing SNPs (minor allele frequency > 0.01) or indels from the 1000 Genomes project were removed from the data. SNP and indel mapping was conducted by Dr Mina Ryten and Dr Adaikalavan Ramasamy (Department of Medical and Molecular Genetics, King's College London). Cross-hybridizing probes, and probe sets containing less than three probes, were also excluded. Detection above background (DABG) noise was calculated for all .CEL files and probe sets were filtered using the Affymetrix Power Tools (Affymetrix) `-a dabg` command. This command calculates the probability that a probe set is expressed above background noise. Probe sets with DABG p-values > 0.01 in 50% of resting or IFN-α-stimulated samples were filtered out. Probes and probe sets that failed QC filters were removed from the data using the statistical package R and the `--kill-list` and `--probeset-ids` commands in Affymetrix Power Tools. Using data .CEL files and the final list of probe sets that passed QC, intensity signals were normalized at exon-level and $\log_2$-transformed using the robust multi-array average (RMA) algorithm in the Affymetrix Expression Console software, build 1.2.1.20 (Affymetrix). Array hybridization quality was verified using Affymetrix Expression Console according to the recommendations of the Affymetrix Quality Assessment of Exon and Gene Arrays White Paper (http://media.affymetrix.com/support/technical/whitepapers/exon_gene_arrays_qa_whitepaper.pdf), including visual inspection of microarray chips, intensity signal histogram and boxplots, and detection of positive and negative controls.

2.4.2 Replicate sample correlation

Duplicate data for COX, QBL, PGF, and DBB available from a different batch were used as technical replicates. For the twins' *ex vivo* B cells, duplicate data for one monozygotic twin pair processed in both study batches were used as technical replicates. Same-batch sibling data from the twin pair were used as biological replicates. Correlation between replicates was assessed using a Spearman correlation test in R, using the following command:

```
> cor(data,use="all.obs",method="spearman").
```

2.4.3 Principal components analysis

Principal Components Analysis (PCA) of the twins' *ex vivo* B cell and MHC-homozygous LCL datasets was performed in R, using the following commands:

```
> fit <- princomp(data, cor=TRUE)
> loadings(fit)
> plot(fit,type="lines")
```

Data structure was visualized in a PCA plot using Partek Genomic Suite version 6.6 (Partek Incorporated), and sample outliers were removed from the data sets. A total of 81 samples from 49 individual twins, and all 11 MHC-homozygous LCL samples were put forward for analysis.

2.5 Exon array data normalization

Intensity signals from exon array probe sets and samples that passed QC filters were normalized at exon level (i.e. probe set level), and gene level (i.e. transcript cluster level) in Partek Genomic Suite (Partek GS). Exon-level data were generated by normalization of probe data using the RMA algorithm and GC background correction. Probe sets were summarized to generate gene-level data by calculating the winsorized mean (10% and 90%), which reduces the effect of outlier probe sets. Both exon-level and gene-level summarization result in $\log_2$-transformed data. Batch effects were accounted for in exon- and gene-level data sets using the batch-remove algorithm in Partek GS. Batch-removal was validated in R using the ComBat function (Leek et al., 2012).

2.6 Data analysis

2.6.1 Differential gene expression

Differential gene expression induced by IFN-α and genotypes was calculated for MHC-homozygous LCL and twins' *ex vivo* B cell data sets using a mixed-model analysis of variance (ANOVA). Analysis were performed using Partek GS. For the analysis of *ex vivo* B cell data, separate models were used to calculate the effect of IFN-α stimulation (equation 1), the effect of

genotype at the cell resting state (equation 2), and the effect of genotype in IFN-α-stimulated cells (equation 3).

(1) $Y = \mu + \text{treatment} + \text{individual ID} + \text{twin ID} + PC1 + PC3 + \text{error}$

(2) $Y_{resting} = \mu + \text{genotype} + \text{treatment} + \text{individual ID} + \text{twin ID} + PC1 + PC3 + \text{error}$

(3) $Y_{IFN\text{-}\alpha} = \mu + \text{genotype} + \text{twin ID} + PC1 + PC3 + \text{error}$

The fitted ANOVA models regressed expression levels at each gene (Y) on fixed-effect terms (treatment and/or genotype), and on random-effect terms denoting individual ID, family structure and zygosity (twin ID), and principal components (PCs) explaining most of the data variability (PC1 and PC3, as shown in Chapter 4, section 4.2). μ represents the intercept. The model testing genotypic effects at cell resting state (equation 2) was applied to all samples, and adjusted for the effect of IFN-α treatment. The model testing genotypic effects at IFN-α-stimulated state (equation 3) was applied exclusively to IFN-α-stimulated samples.

The ANOVA model fitted for the MHC-homozygous LCL data set is shown in equation 4.

(4) $Y = \mu + \text{genotype} + \text{cell line ID} + \text{error}$

The model regressed expression levels at each gene (Y) on genotype and cell line ID (random-effect), in order to adjust for the effect of duplicate samples. μ represents the intercept. All p-values were corrected for multiple testing using the false discovery rate (FDR) method.

2.6.2 Gene-environment interaction

The interaction between IFN-α and *DRB1*03:01* was tested in the twins' *ex vivo* B cell gene-level data set using Partek GS. To test whether genotypic effects change across different treatment conditions, the interaction term treatment × genotype was included in the ANOVA model, in addition to treatment, genotype, and random-effect covariates (equation 5).

(5) $Y = \mu$ + treatment + genotype + (treatment × genotype) + individual ID + twin ID + PC1 + PC3 + + error

Y represents the gene expression trait and μ represents the intercept. Genes that are affected by one variable (genotype) differentially across a second variable (treatment condition) will have small p-values. All p-values were corrected for multiple testing using the FDR method.

2.6.3 Alternative splicing

Alternative splicing between samples of different treatment groups or genotypes was calculated in the twins' *ex vivo* B cell exon-level data set using Partek GS's alternative splice ANOVA. This model tests the interaction between the factor of interest and exon (factor × exon). Separate models were used to calculate the effect of IFN-α stimulation (equation 6), the effect of genotype at the cell resting state (equation 7), and the effect of genotype in IFN-α-stimulated cells (equation 8).

(6) $Y = \mu$ + sample ID + exon + treatment + (treatment × exon) + individual ID + twin ID + error

(7) $Y_{resting} = \mu$ + sample ID + exon + genotype + (genotype × exon) + treatment + individual ID + twin ID + error

(8) $Y_{IFN\text{-}\alpha} = \mu$ + sample ID + exon + genotype + (genotype × exon) + twin ID + error

Y represents the exon expression trait, μ is the intercept, and sample ID is a sample-to-sample effect. Each model adjusted for the same covariates as those used in the standard ANOVA models (equations 1, 2, and 3, respectively); PC1 and PC3 were the exception, due to confounding with the factor × exon interaction terms. The splice ANOVA model testing genotypic effects at cell resting state (equation 7) was applied to all samples, and adjusted for the effect of IFN-α treatment. The model testing genotypic effects at IFN-α-stimulated state (equation 8) was applied exclusively to IFN-α-stimulated samples. The resulting p-value for each gene shows the probability that one or more exons are differentially expressed between treatment groups or genotypes. As alternative splicing events generally assume no changes in overall transcript expression levels, genes showing differential expression at $P < 0.05$ (gene-level) were removed from the analysis. All p-values were corrected for multiple testing using the FDR method.

2.6.4 Ingenuity Pathway Analysis

Pathway and disease enrichment in genes showing differential expression or alternative splicing in the twins' *ex vivo* B cell data set were calculated in Ingenuity Pathway Analysis (Qiagen), using a right-tailed Fisher's Exact Test. The 2 × 2 contingency table and equation used to compute enrichment p-values are shown below.

	DR* genes	Non-DR genes	Row total
Molecules in pathway	a	b	a + b
Molecules not in pathway	c	d	c + d
Column total	a + c	b + d	n

* DR – Differentially regulated

The probability of association between genes differentially regulated by IFN-α and a pathway or disease gene set is:

$$p = \frac{\binom{a+b}{a}\binom{c+d}{c}}{\binom{n}{a+c}} = \frac{(a+b)!\ (c+d)!\ (a+c)!\ (b+d)!}{a!\ \ b!\ \ c!\ \ d!\ \ n!}$$

2.6.5 eQTL analysis

eQTL analysis was performed using the Matrix eQTL R package (Shabalin, 2012). The correlation between SNP genotype and probe set expression levels was calculated using a linear regression model with additive genotype effects (Y = μ + genotype + covariates + error, where Y is the gene expression trait, and μ is the intercept). eQTLs were defined as *cis* if the genomic distance between the SNP and the probe set start site was less than or equal to 2.5 MB. eQTLs were defined as *trans* if the distance between the SNP and probe set was greater than 2.5 MB, or if they were located on different chromosomes. All p-values were adjusted for multiple testing using the FDR correction method. The analyses were conducted for i) the *ex*

vivo B cell data set from Fairfax and colleagues, adjusting for 33 PCs as recommended by the authors (Fairfax et al., 2012); ii) the HapMap LCL data set; and iii) the twins' *ex vivo* B cell gene-level data set generated from exon arrays (sections 2.3, 2.4, and 2.5), adjusting for covariates as described in section 2.6.1. For the MuTHER LCL data set, *cis*-eQTL data were obtained through the Genevar resource, and *trans*-eQTL data were obtained from Dr Kerrin Small (Department of Twin Research and Genetic Epidemiology, King's College London; personal communication). SNP genotype data were available for all cohorts studied.

2.7 TaqMan qPCR

TaqMan quantitative polymerase chain reaction (qPCR) was used for validation of results obtained from array experiments, including the effects of IFN-α treatment on B cell gene expression, and eQTLs within *DRB1*03:01* haplotypes. The principle of qPCR is based on the detection and quantification of a fluorescence signal that emanates during the exponential stages of PCR amplification. TaqMan Gene Expression Assays contain a pair of unlabelled PCR primers, and a TaqMan probe linked to a fluorophore and a quencher. When the probe is intact, the quencher inhibits any fluorescence signal emitted by the fluorophore. Quantification of gene expression relies on the 5´-3´ exonuclease activity of the *Taq* polymerase to cleave the dual-labelled probe during PCR amplification of the target. The resulting light emission is proportional to the amount of PCR product generated, thus providing a quantitative measurement. The presence of a probe, in addition to PCR primers, greatly increases target specificity of TaqMan qPCR compared to other qPCR methods.

Twenty-four pre-designed TaqMan Gene Expression Assays (Life Technologies) were obtained for qPCR analysis, according to pre-defined criteria outlined in Chapters 4 and 5. Target genes included type 1 IFN signature genes, idiopathic and monogenic SLE-associated genes, and genes with evidence of differential expression in *DRB1*03:01* haplotypes. TaqMan assays and their corresponding target transcripts are listed in Table 2-1. TaqMan qPCRs were prepared using 4 µl cDNA from twins' *ex vivo* B cell samples, 0.5 µl of TaqMan assay, 5 µl of TaqMan Universal PCR Master Mix (Life Technologies), and 0.5 µl H_2O per reaction. qPCR was performed in triplicate using the 7900HT Fast Real-Time PCR System (Applied Biosystems).

The cycle conditions were as follows: 50°C for 2 min, 95°C for 10 min, followed by 40 cycles of 95°C for 15 seconds (sec) and 60°C for 1 min.

Table 2-1 TaqMan Gene Expression Assays used for qPCR validation experiments.

	Gene	**TaqMan Assay**
qPCR internal control genes	*UBB*	Hs00430290_m1
	TBP	Hs00427620_m1
	OAZ1	Hs00427923_m1
Type 1 IFN signature genes	*IFIT1*	Hs00356631_g1
	IFI44	Hs00951349_m1
	IFI44L	Hs00915292_m1
	ISG15	Hs00192713_m1
	IRF7	Hs01014809_g1
Idiopathic SLE susceptibility genes	*IRF5*	Hs00158114_m1
	IFIH1	Hs01070332_m1
	IKZF1	Hs00958474_m1
	ITGAM	Hs00355885_m1
	TNPO3	Hs01110576_m1
	FCGR2A	Hs01017702_g1
	PTPN22	Hs01587519_g1
	FCGR3B	Hs04334165_m1
Monogenic SLE susceptibility genes	*SAMHD1*	Hs01122750_m1
	DNASE1L3	Hs00172840_m1
	PRKCD	Hs01090040_m1
	RNASEH2C	Hs00611050_g1
	ACP5	Hs00356261_m1
Genes dysregulated in *HLA-DRB1*03:01* haplotypes	*BTN3A2*	Hs00389328_m1
	PSMB9	Hs00160610_m1
	HLA-DPB1	Hs03045105_m1

Absolute threshold cycle (Ct) values and reaction slopes were obtained from the SDS version 2.4 software (Applied Biosystems). Within triplicate samples, those showing standard deviations over 0.3 were considered to be of low reproducibility, and were therefore removed from the analysis. Reaction efficiencies were calculated for each gene ($E = 10^{-1/slope}$). The mean Ct values of genes of interest were normalized to the internal control genes *UBB*, *TBP*, and *OAZ1*, using $\Delta Ct = Ct_{gene} - Ct_{controls}$. The resulting ΔCt values were used to calculate differential gene expression. All three control genes have previously been reported to be suitable qPCR controls (Hruz et al., 2011, Ersahin et al., 2014), and showed similar expression levels across all treatment and genotype groups in the exon array data.

2.8 Western blot

The effect of the *DRB1*03:01* haplotype on *BTN3A1*, *BTN3A2*, and *BTN3A3* protein expression was assessed by Western blot analysis. Two *DRB1*03:01*-homozygote LCLs (COX and QBL) and three non-*DRB1*03:01* LCLs (MANN, APD, and PGF) were used. LCLs were thawed and cultured as described in section 2.3.5. Eight million cells were harvested at resting state. Cells were lysed with 1 ml cold RIPA buffer (Thermo Scientific), rotated for 30 min at 4 °C and sonicated 3 times for 5 sec. Lysates were centrifuged for 20 min at 12,000 × g, at 4 °C, and protein was quantified using the Bio-Rad DC Protein Assay (Bio-Rad), according to the manufacturer's instructions.

Lysates were denatured for 10 min at 95 °C in 4 × loading buffer. Samples were loaded on a 10% Mini-PROTEAN TGX Precast Protein Gel (Bio-Rad), and separation by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) was conducted at 200 volts for 45 min in running buffer (250 mM Tris, 1.92 M glycine, 1% SDS, Fisher Scientific). Protein was transferred to a nitrocellulose membrane (GE Healthcare) at 100 volts for 30 min in transfer buffer (running buffer with 20% methanol), and the membrane was saturated with 5% BSA/PBS for 1 h. Membranes were incubated with primary antibodies in 1% BSA/PBS/0.05% Tween (Sigma-Aldrich) overnight at 4 °C. The primary antibodies used were 1.1 µg/ml anti-BTN3A2 (28010002, Bio-Techne), 0.9 µg/ml anti-BTN3A3 (HPA007904, Atlas Antibodies), 1 µg/ml anti-CD277 (14–2779-82, eBioscience), and 0.05 µg/ml anti-beta-actin (sc-47778, Santa Cruz Biotechnology).

Membranes were washed 3 times with PBS/0.1% Tween for 5 min and incubated with horseradish peroxidase (HRP)-conjugated secondary antibody (0.2 µg/ml anti-mouse, W4021, Promega, or 0.23 µg/ml anti-rabbit, sc-2030, Santa Cruz Biotechnology) in 1% BSA/PBS/0.05% Tween for 1 h at RT. Membranes were again washed 3 times with PBS/0.1% Tween and incubated with the Luminata Crescendo Western HRP substrate (Millipore) for 3 min. Following membrane exposure to light-sensitive films for 30 sec or 1 min, films were developed using the Compact X4 film processor (Xograph Imaging Systems).

Protein intensities were quantified using the ImageJ software (National Institutes of Health). Data were normalized to beta-actin, and differential expression was calculated using a Student's *t*-test. Experiments were done in duplicate.

2.9 Bioinformatic analysis

Information on gene structure and function was explored using the Ensembl browser (http://www.ensembl.org/index.html), BioGPS (http://biogps.org/#goto=welcome), and the NCBI Entrez Gene browser (http://www.ncbi.nlm.nih.gov/gene).

2.9.1 Linkage disequilibrium

Linkage disequilibrium was calculated using Haploview version 4.2 (Broad Institute) for different 1000 Genomes populations, and for all cohorts described in this chapter. Tag SNPs for disease-associated variants were obtained using the Tagger algorithm in Haploview ($r^2 > 0.80$).

2.9.2 SNP functional annotations

The HaploReg version 4.1 resource was used to access summary data for published eQTLs within SLE-associated haplotypes of interest (www.broadinstitute.org/mammals/haploreg, accessed September 2015). Regulatory features overlapping the SLE-protective *MSH5* haplotype were obtained from the ENCODE database, using the UCSC browser (http://genome.ucsc.edu/) and ASSIMILATOR (http://assimilator.mhs.manchester.ac.uk/cgi-bin/assimilator.pl, accessed June 2013). Regulatory sites of interest included experimentally-validated transcription factor binding sites (TFBS), microRNA binding sites, DNA

hypersensitivity sites, and methylated DNA. Predicted TFBS disruption by SNPs of interest was calculated using JASPAR (http://jaspar.genereg.net/, accessed June 2013). The effect of non-synonymous SNPs on protein function was predicted using SIFT (http://sift.jcvi.org/www/SIFT_dbSNP.html) and Polyphen-2 version 2 (http://genetics.bwh.harvard.edu/ pph2/).

Chapter 3 - eQTL analysis of MHC haplotypes associated with SLE using published data sets

3.1 Introduction

3.1.1 Genetically independent associations at the MHC in SLE

The major histocompatibility complex (MHC) confers the strongest genetic risk for systemic lupus erythematosus (SLE) in all populations studied to date, with the most significant associations consistently arising from haplotypes harbouring the classical MHC class II genes *HLA-DRB1*, *HLA-DQA1*, and *HLA-DQB1* (Tsao, 2004, Fernando et al., 2012, International MHC and Autoimmunity Genetics Network et al., 2009). The causal variants underlying MHC associations have remained elusive, owing to the extent of linkage disequilibrium (LD) across the region, the high polymorphism of classical human leukocyte antigen (HLA) genes, and copy number variation, as well as the heterogeneity of disease phenotype. The strongest associations with SLE risk in Europeans arise from the *DRB1* alleles **03:01* (odds ratio, OR = 2.10-2.70), **15:01* (OR = 1.33-1.60), and **08:01* (OR = 3.50), and their respective haplotypes (Graham et al., 2002, Uribe et al., 2004, International MHC and Autoimmunity Genetics Network et al., 2009, Fernando et al., 2012, Morris et al., 2012).

The *DRB1*03:01* risk allele tags an extended ancestral haplotype which is common in European populations (frequency ~ 10%). Association studies in European-derived populations, such as Mexican, Tunisian, and Hispanic, have also shown an increased frequency of *DRB1*03:01* in SLE patients, compared to matched controls. However, in non-European populations, association of *DRB1*03:01* with SLE have been inconsistent (Smikle et al., 2002, Ayed et al., 2004, Cortes et al., 2004, Uribe et al., 2004, Kim et al., 2009). The association of *DRB1*03:01* with autoantibody subsets in SLE supports a direct role for the *DRB1*03:01* allele itself in disease pathogenesis, given that MHC class II presentation to T cells is required for the initiation of antibody-mediated immunity. Furthermore, the association of the *DRB1*03:01* haplotype with several other autoimmune diseases (as previously shown in Chapter 1, Table 1-2) reiterates the importance of this haplotype in the pathogenesis of antibody-mediated diseases.

The association of *DRB1*15* with SLE across different populations reflects the geographical distribution of *DRB1*15* allele frequencies. *DRB1*15:01* is associated with SLE in European, Korean, Japanese, Taiwanese, and Thai populations; *DRB1*15:02* is associated in Filipino and Taiwanese populations; and *DRB1*15:03* is associated in African-derived populations (Fernando et al., 2008, Fernando et al., 2012, Cortes et al., 2004, Ayed et al., 2004, Bang et al., 2015, Shimane et al., 2013, Lu et al., 1997, Sirikong et al., 2002, Suggs et al., 2011, Uribe et al., 2004). A trans-ancestral study using European and South East Asian populations successfully fine-mapped the *DRB1*15* association to an 87 kilobases (KB) interval encompassing *DRB1* and *DQA1* (Fernando et al., 2012). In contrast to *DRB1*03:01* and *DRB1*15:01*, the *DRB1*08:01* allele has only shown association with SLE in some cohorts of Southern European ancestry (Graham et al., 2002, Uribe et al., 2004, Fernando et al., 2012).

The prevailing hypothesis explaining the association of MHC haplotypes with autoimmunity risk proposes that amino acid variation resulting from the risk-associated HLA alleles disrupt central tolerance, by predisposing to aberrant HLA-T cell receptor (TCR) interactions. This may in turn affect the negative selection of autoreactive CD4+ T cells in the thymus and, upon encountering self-antigens in the periphery, result in the activation of autoantibody-mediated immunity. An additional hypothesis suggests that variants within disease-associated MHC haplotypes also contribute to disease risk through mechanisms other than HLA amino acid variation. Furthermore, there is increasing evidence to suggest that SLE associations at the MHC result from the combinatorial effects of multiple independent HLA and/or non-HLA variants (Morris et al., 2012). Recent case-control studies have successfully fine-mapped a number of independent loci at the MHC, by combining dense single nucleotide polymorphism (SNP) genotyping, HLA typing, and the use of multi-ethnic cohorts. A recent study by Fernando and colleagues conducted a trans-ancestral mapping analysis in SLE using cohorts of British (United Kingdom, U.K.), Spanish, and Filipino ancestry (Fernando et al., 2012). The different LD structure exhibited by the three populations enabled fine-mapping of MHC associations to *HLA-G*, *HLA-B/HLA-C*, and *PSORSC1* in class I; *HLA-DPB1* in class II; and *MSH5* in class III. Interestingly, both SLE-risk and SLE-protective variants were found within the *MSH5* gene. The risk variant, rs3131379, showed association in the U.K. and Spanish cohorts, and is in moderate LD with *DRB1*03:01* (r^2 = 0.62, U.K. cohort, and r^2 = 0.28, Spanish cohort). The protective SNP,

rs409558, showed an independent association in the Spanish and Filipino cohorts, and was fine-mapped to a 20 KB haplotype encompassing most of the *MSH5* gene. Interestingly, variants within *MSH5* have been associated with other autoimmune diseases, including rheumatoid arthritis, type 1 diabetes, and multiple sclerosis, as well as immunoglobulin A (IgA) deficiency, common variable immunodeficiency, and premature ovarian failure (Sirota et al., 2009, Sekine et al., 2007, Mandon-Pepin et al., 2008). Independent associations of *DPB1*, a classical class II gene, with SLE have been reported in Spanish (rs3117213, risk), Filipino (rs2071351, protective), and African-American (rs2071349, risk) populations (Fernando et al., 2012, Ruiz-Narvaez et al., 2011). The Spanish risk variant is also associated with anti-citrullinated protein antibody (ACPA)-positive rheumatoid arthritis, whilst other SNPs within *DPB1* have shown association with several autoimmune diseases, such as Graves' disease, systemic sclerosis, and multiple sclerosis (Ding et al., 2009, Takahashi et al., 2006, Zhou et al., 2009, Wu et al., 2009a). In conclusion, fine-mapping SLE association signals arising within the MHC has had some success, as outlined above. However, the causal variants underlying these associations, and their functional consequences, are yet to be determined.

3.1.2 eQTLs arising within the MHC region

Most variants associated with complex diseases in genome-wide association studies (GWAS) are located in non-coding regions of the genome (Hindorff et al., 2009). There is increasing evidence suggesting that non-coding variants alter susceptibility to disease by regulating gene expression or alternative splicing. For example, SLE-associated variants within the *BLK* and *STAT4* genes correlate with altered transcript levels in B cells and peripheral blood mononuclear cells (PBMCs) from healthy individuals (Hom et al., 2008, Abelson et al., 2009). Given the relatively large effect sizes of *DRB1* haplotype associations with SLE and other autoimmune diseases, it is possible that regulatory variants within these haplotypes may contribute to the haplotypic disease risk, in addition to HLA amino acid variation. One example is the *cis*-expression quantitative trait locus (eQTL) identified for the *TNF* promoter variant, -308/G>A, which is present in *DRB1*03:01* haplotypes (Wilson et al., 1993, Helmig et al., 2011, Abraham and Kroeger, 1999).

In order to investigate the functional impact of SLE-associated MHC haplotypes further, one may hypothesize that eQTLs within risk and protective MHC haplotypes contribute to disease

pathogenesis by altering gene expression. Recent studies have shown that carrying specific HLA alleles can affect gene expression MHC-wide (i.e. *in cis*, Vandiedonck et al., 2011) and outside the MHC (*in trans*, Fairfax et al., 2012). However, only a limited number of studies have investigated gene expression patterns associated with SLE-risk MHC haplotypes. An MHC-wide study by Vandiedonck and colleagues using lymphoblastoid cell lines (LCLs) demonstrated differential expression of MHC genes between two *DRB1*03:01*-homozygous LCLs (COX and QBL) and one *DRB1*15:01*-homozygous LCL (PGF, Vandiedonck et al., 2011). To date, there has not been a comprehensive investigation of *cis*- and *trans*-eQTLs within SLE-associated MHC haplotypes in large data sets.

3.1.3 Aims and study design

In order to gain further insight into the mechanisms underlying disease associations, the effect of SLE-associated MHC haplotypes on gene expression was investigated in healthy individuals, using published data sets. The eQTL analyses were conducted using data derived from *ex vivo* B cells and LCLs (Epstein-Barr virus (EBV)-transformed B cells), which were selected due to the key role of B cells in the pathogenesis of SLE, and the enrichment of SLE-associated variants in B cell-specific *cis*-eQTLs (Fairfax et al., 2012). *Cis*- and *trans*-eQTLs were investigated for *DRB1*, *DPB1*, and *MSH5* haplotypes associated with SLE risk and protection. The potential functional consequences of variants within the *MSH5* protective haplotype were then explored using bioinformatic tools, in an attempt to further elucidate the functional role of this association.

eQTL analyses were conducted using tag SNPs for the European *DRB1*03:01* (rs2187668, intronic) and *DRB1*15:01* risk haplotypes (rs3135391, exonic), the Spanish *DPB1* risk haplotype (rs3117213, intergenic), the Filipino *DPB1* protective haplotype (rs2071351, intronic), the African-American *DPB1* risk haplotype (rs2071349, intronic), and the Spanish *MSH5* risk (rs3131379, intronic) and protective (rs409558, intronic) haplotypes. Three data sets derived from *ex vivo* B cells and LCLs of Northern European ancestry were used in this study: (i) LCLs from the HapMap CEU population (sample size, N = 60, Stranger et al., 2007), (ii) LCLs from the Multiple Tissue Human Expression Resource (MuTHER, N = 777, Grundberg et al., 2012), and (iii) *ex vivo* B cells from the study by Fairfax and colleagues (N = 281, Fairfax et al., 2012). eQTL analyses were performed using the HapMap LCL and the *ex vivo* B cell data sets. Available eQTL data from MuTHER LCLs were accessed from the GeneVar resource.

Genotype counts for each data set are detailed in Appendix A. In order to corroborate results, a search of the HaploReg resource for published *cis*-eQTLs in various cell types was carried out. Given the extent of LD at the MHC, *cis*-eQTLs for the analyses of HapMap LCLs and *ex vivo* B cells were defined as any association between SNPs and probes located at a distance of up to 2.5 megabases (MB). All other data sets had pre-determined *cis*-eQTLs as associations within 1 MB of the query SNP. The significance threshold in all published data sets was pre-set at a false discovery rate (FDR) of 5%. Only raw p-values are presented in this chapter, for comparison with published data.

Results

3.2 The *MSH5* protective haplotype

3.2.1 *Cis*-eQTLs arising from the *MSH5* protective haplotype

In order to investigate the potential regulatory effects of the *MSH5* protective haplotype, the SLE-associated SNP rs409558 was used to query the HapMap and MuTHER LCL data sets. As genotype data for rs409558 was not available in the *ex vivo* B cell data set from Fairfax and colleagues, rs2293861, a proxy SNP for rs409558, was queried instead (r^2 = 1.00 in Northern European populations from 1000 Genomes).

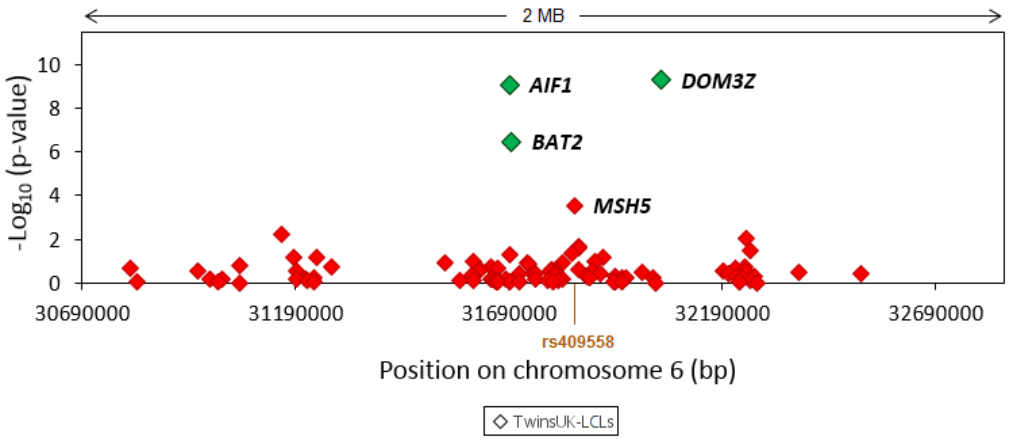


Figure 3-1 Regional plot of *cis*-eQTL associations for the SLE-protective *MSH5* SNP, rs409558, in LCLs from the MuTHER cohort (Grundberg et al., 2012).

The strength of association with expression probes (lozenges) located within 1 MB of the SNP is plotted as -$\log_{10}$(p-value) against genomic location. *Cis*-associations significant at an FDR of 5% are coloured in green.

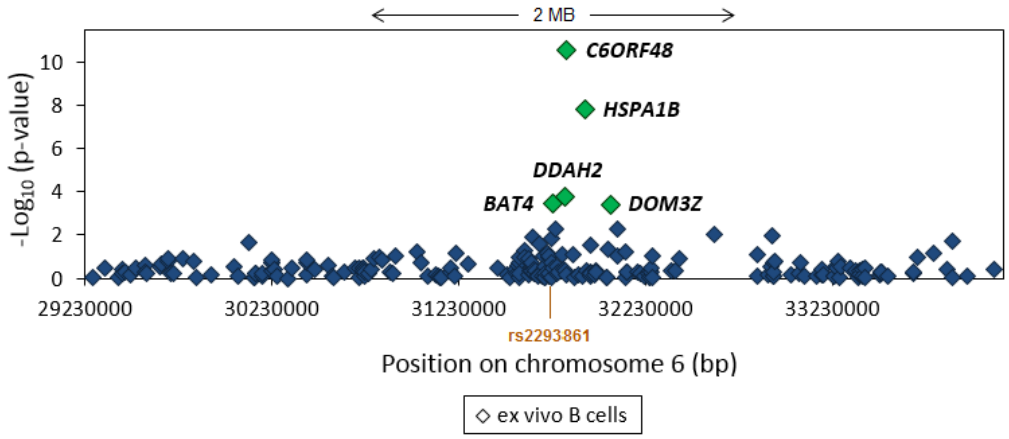


Figure 3-2 Regional plot of *cis*-eQTL associations for the SLE-protective *MSH5* SNP, rs2293861, in *ex vivo* B cells from Fairfax et al. (2012).

The strength of association with expression probes (lozenges) located within 2.5 MB of the SNP is plotted as $-\log_{10}$(p-value) against genomic location. The arrows indicate the region within 1 MB of the SNP. *Cis*-associations significant at an FDR of 5% are coloured in green.

Cis-eQTL analysis for rs409558 in the MuTHER LCLs showed significant associations with three transcripts: *DOM3Z* ($P = 4.85 \times 10^{-10}$), *AIF1* ($P = 8.72 \times 10^{-10}$), and *BAT2* ($P = 3.46 \times 10^{-7}$, Figure 3-1). In contrast, analysis of HapMap LCLs showed no significant *cis*-eQTLs. However, one of five available *MSH5* probes, ILMN_1810834, was differentially expressed prior to multiple test correction in both HapMap LCLs ($P = 2.45 \times 10^{-3}$, fold change (FC) = 1.13, FDR > 5%) and MuTHER LCLs ($P = 2.98 \times 10^{-4}$, FC = 1.04, FDR > 5%). This probe targets the 3' end of *MSH5* transcripts, including a read-through transcript to the downstream gene at the 3' end, *SAPCD1* (Figure 3-3). Similarly, *cis*-association with a probe uniquely targeting *SAPCD1* also showed suggestive association in the MuTHER LCLs (ILMN_2134765, $P = 4.69 \times 10^{-2}$, FC = 1.03, FDR > 5%). In *ex vivo* B cells, rs2293861 was a significant eQTL for five genes (*C6orf48*, *DDAH2*, *HSPA1B*, *BAT4*, and *DOM3Z*), all of which are located within 125 KB of the *MSH5* locus (Figure 3-2). Again, the *MSH5* probe ILMN_1810834 showed association prior to correcting for multiple testing ($P = 1.43 \times 10^{-2}$, FC = -1.04, FDR > 5%). Despite contradicting directions of effect, it is of interest that the p-values for this probe fell just under the significance threshold in all three data sets. Conversely, none of the other *MSH5* or *SAPCD1* probes present in the data sets showed suggestive *cis*-associations (P > 0.12).

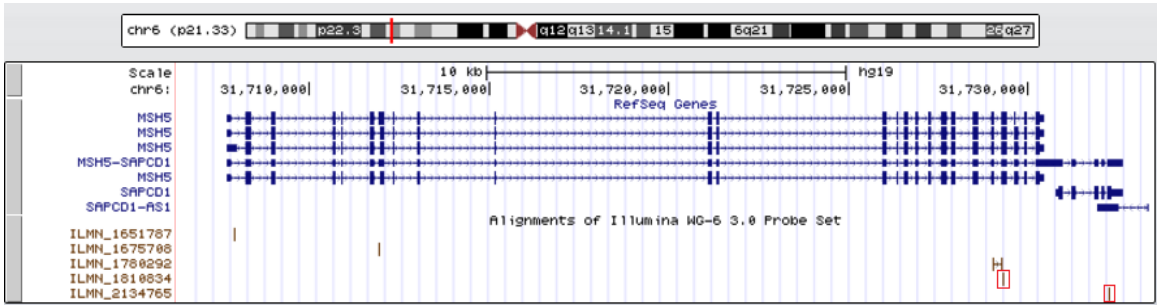


Figure 3-3 UCSC Genome Browser screenshot showing target locations for *MSH5* and *SAPCD1* probes from the Illumina Sentrix Human-6 BeadChip array in the reference MHC sequence.

All probes shown are shared by the Illumina platforms used for the HapMap LCL data set (Stranger et al., 2007), the MuTHER LCL data set (Grundberg et al., 2012), and the *ex vivo* B cell data set (Fairfax et al., 2012). The diagram illustrates *MSH5* and *SAPCD1* transcripts from RefSeq gene annotations (top) and Sentrix Human-6 BeadChip array probes (bottom). The probes that showed suggestive associations with the SLE-protective haplotype tagged by rs409558 in one or more data sets are outlined in red ($P < 0.05$, FDR > 5%).

The most significant *cis*-association with rs409558 in *ex vivo* B cells was to *C6orf48* ($P = 2.83 \times 10^{-11}$, FC = -1.06, Figure 3-4). A search using the HaploReg resource showed that the *C6orf48* *cis*-association was validated in *ex vivo* naïve monocytes ($P = 1.34 \times 10^{-21}$, Fairfax et al., 2014) and whole blood from two independent studies ($P = 2.20 \times 10^{-7}$, GTEx Consortium, 2015, and $P = 2.37 \times 10^{-15}$, Westra et al., 2013). Additionally, *C6orf48* was also a significant *cis*-gene in skin samples from the MuTHER cohort ($P = 6.92 \times 10^{-4}$) and the GTEx Consortium ($P = 3.4 \times 10^{-7}$, GTEx Consortium, 2015), as well as in tibial nerve samples ($P = 7.2 \times 10^{-8}$, GTEx Consortium, 2015) and testis samples ($P = 1.50 \times 10^{-6}$, GTEx Consortium, 2015). However, *C6orf48* was not a significant *cis*-gene in the MuTHER LCLs ($P = 7.21 \times 10^{-2}$) and failed quality control (QC) in the HapMap LCL data set. The *C6orf48* locus harbours two small nucleolar ribonucleic acid (snoRNAs) encoded within intron 1, *SNORD48* and *SNORD52*. Accordingly, it was assessed whether rs409558 was also an eQTL for the two snoRNAs in LCLs and *ex vivo* B cells. The *SNORD52* probe did not show a significant association with genotype, while the *SNORD48* probe failed QC in all data sets.

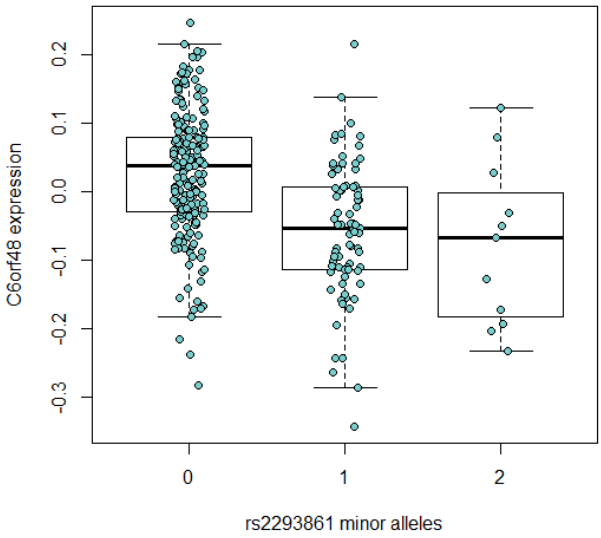


Figure 3-4 eQTL box plot for the *MSH5* protective haplotype tag SNP, rs2293861, and *C6orf48* expression in *ex vivo* B cells from Fairfax et al. (2012).

rs2293861 is a significant *cis*-eQTL for *C6orf48* ($P = 2.83 \times 10^{-11}$, FDR < 5%). The box plot shows median, upper, and lower quartiles of expression levels for each genotype group. Whiskers show the maximal and minimal non-outlying data points.

The *DOM3Z cis*-association observed in MuTHER LCLs and *ex vivo* B cells was replicated in naïve monocytes ($P = 1.46 \times 10^{-6}$, Fairfax et al., 2014), whole blood ($P = 5.21 \times 10^{-13}$, Westra et al., 2013), and trended towards significance in HapMap LCLs ($P = 2.41 \times 10^{-2}$, FDR > 5%). *HSPA1B* and *BAT4* were the only other *cis*-genes in *ex vivo* B cells that were validated in independent cohorts (LCLs from Lappalainen et al., 2013, monocytes from Fairfax et al., 2014, and whole blood from Westra et al., 2013). Table 3-1 summarises all *cis*-eQTL results that were replicated, or showed suggestive evidence of replication, in published data sets from LCLs, *ex vivo* B cells, *ex vivo* monocytes, and whole blood. Interestingly, not all *cis*-genes showed the same direction of effect in the different cell populations; for example, *DOM3Z* was down-regulated in *ex vivo* B cells but up-regulated in MuTHER LCLs.

Table 3-1 Summary of *cis*-eQTLs for rs409558 that were replicated in independent data sets from *ex vivo* B cells, LCLs, *ex vivo* monocytes, and whole blood.

Correlated transcript	***Ex vivo* B cells (Fairfax et al., 2012)‡**	**MuTHER LCLs (Grundberg et al., 2012)§‡**	**HapMap LCLs (Stranger et al., 2007)‡**	**LCLs (Lappalainen et al., 2013)*†**	***Ex vivo* naïve monocytes (Fairfax et al., 2014)*‡**	**Whole blood (Westra et al., 2013)*‡**	**Whole blood (GTEx Consortium, 2015)*§†**
	N = 281	N = 777	N = 60	N = 462	N = 421	N = 5,311	N = 156
C6orf48	✓ P = 2.83×10^{-11} FC = -1.06	*P = 7.21×10^{-2}* *FC = 1.02*	ND	ND	✓ P = 1.34×10^{-21} ND	✓ P = 2.37×10^{-15} ND	✓ P = 2.20×10^{-7} FC = -1.12
DOM3Z	✓ P = 4.18×10^{-4} FC = -1.07	✓ P = 4.85×10^{-10} FC = 1.09	*P = 2.41×10^{-2}* *FC = -1.11*	ND	✓ P = 1.46×10^{-6} ND	✓ P = 5.21×10^{-13} ND	ND
HSPA1B	✓ P = 1.51×10^{-8} FC = -1.15	ND	ND	ND	✓ P = 3.85×10^{-6} ND	✓ P = 1.04×10^{-59} ND	ND
BAT4	✓ P = 3.25×10^{-4} FC = 1.07	ND	*P = 7.24×10^{-3}* *FC = 1.08*	✓ P = 5.11×10^{-14} ND	✓ P = 3.28×10^{-4} ND	ND	ND

The significance threshold was set to an FDR of 5% in all data sets. Italicized p-values and fold changes indicate suggestive associations (i.e. significant at P < 0.05 prior to correction for multiple testing, but non-significant at an FDR of 5%). All cohorts included healthy individuals of European ancestry, and data sets were corrected for population structure.
**Cis*-eQTL data accessed from the HaploReg resource.
§ Data sets adjusted for sex.
† Data sets generated using RNA sequencing.
‡ Data sets generated using expression microarrays.
N – sample size, FC – fold change (effect size per minor allele), ND – no data available.

3.2.2 Functional prediction of variants within the *MSH5* protective haplotype

The mechanisms that may be driving the *cis*-eQTLs at the *MSH5* locus were investigated. The size of the haplotype tagged by rs409558 in European populations was first defined, and a search for functional annotations within this haplotype was then performed using publicly-available data bases. To confirm the independence of the *MSH5* protective signal from HLA haplotypes, the LD between rs409558 and HLA alleles was calculated using genotypic data from U.K. and Spanish populations (data obtained from Fernando et al., 2012). rs409558 showed low LD with all typed alleles from *HLA-B*, *HLA-DRB1*, *HLA-DQA1*, and *HLA-DQB1* in both populations ($r^2 < 0.03$). As the SLE association with this haplotype could be due to either rs409558 itself, or a functional SNP in LD with it, the functional annotations of the SNPs within the protective haplotype in the Spanish (IBS) and British (GBR) populations from 1000 Genomes (1000G) were investigated. The haplotype tagged by rs409558 was defined to a 21 KB region flanking the *MHS5* gene in both the IBS and GBR populations, using a $r^2 > 0.80$ in Haploview's Tagger algorithm. The haplotype comprises thirteen SNPs in the IBS population and six SNPs in the GBR population (Figure 3-5, Appendix B). All SNPs are located within introns of *MSH5* or neighbouring genes, except for rs6905572, which is located in exon 3 of *SAPCD1*. rs6905572 results in a non-synonymous change in the SAPCD1 protein, from a proline to leucine at position 99 (Pro99Leu), which was predicted to be non-deleterious by the SIFT and PolyPhen-2 version 2 tools.

A search was conducted using publicly-available data bases in order to determine whether any SNP within the *MSH5* protective haplotype overlapped putative regulatory elements. All SNPs were analysed using the ASSIMILATOR tool, which queries several data bases including ENCODE and OpenChrom. rs409558 and rs4711279 (located upstream of *C6orf48*) were shown to overlap a high number of regulatory features relating to gene expression, including multiple transcription factor binding sites (TFBS), CpG islands, DNAse I hypersensitivity sites, and histone modifications indicative of active promoters. ASIMILATOR results are summarised in Table 3-2. Data obtained from chromatin immunoprecipitation-sequencing (ChIP-seq) for the rs409558 locus showed strong evidence for binding of E2F1 (HeLa-S3 and MC-7 cells, score of 1000), E2F6 (K562 cells, score of 1000 and HeLa-S3 cells, score of 890), and ZFN263 (T-REx-HEK293 cells, score of 199).

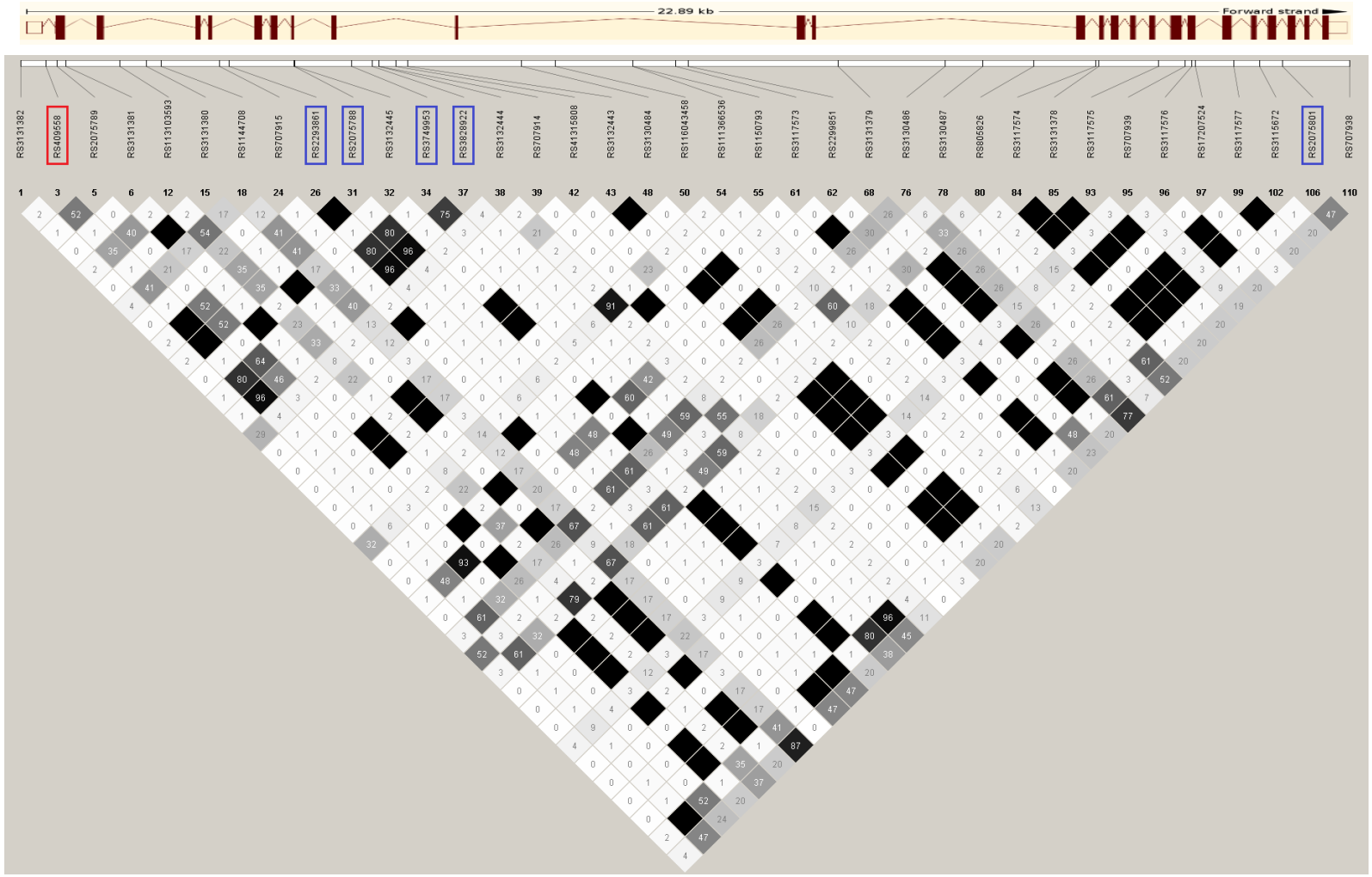


Figure 3-5 LD plot of the *MSH5* locus.

The *MSH5* gene is illustrated at the top (image from Ensembl), and r^2 values between common SNPs (minor allele frequency > 5%) from the GBR 1000 Genomes population are displayed below. The rs409558 variant in intron 1 is highlighted in red, and five proxy SNPs are highlighted in blue (r^2 = 1.00). The LD plot was generated using Haploview version 4.2.

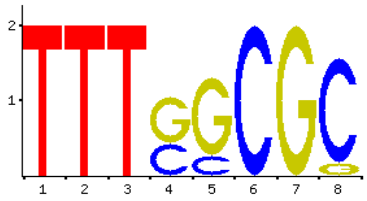
Table 3-2 ASSIMILATOR output of regulatory features overlapping the *MSH5* SLE-protective haplotype.

SNP	Location	CpG Islands	HAIB Methyl450	Histone markers			UNC FAIRE	DNAse I HS			TFBS				Txn Factor ChIP	UW CTCF Binding
				Broad Institute	SYDH	UW		Uniform	Duke	DNase Clusters	Uniform	HAIB	UTA	UC		
rs409558	Intronic	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
rs2293861	Intronic			Yes	Yes	Yes					Yes	Yes	Yes	Yes		Yes
rs2075788	Intronic			Yes	Yes	Yes			Yes			Yes	Yes	Yes		
rs3749953	Intronic			Yes		Yes		Yes	Yes			Yes	Yes			
rs3828922	Intronic			Yes		Yes						Yes				
rs2075801	Intronic			Yes	Yes	Yes						Yes	Yes			
rs6905572	Exonic		Yes	Yes		Yes			Yes		Yes	Yes				
rs9469051	Intronic			Yes		Yes										
rs4711277	Intronic			Yes		Yes										
rs4711279	Promoter		Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
rs116828153	Intronic		Yes	Yes	Yes	Yes		Yes	Yes	Yes	Yes	Yes	Yes		Yes	
rs116201952	3' UTR			Yes	Yes	Yes		Yes			Yes	Yes	Yes		Yes	Yes
rs80260075	Intergenic			Yes	Yes	Yes										

HAIB Methyl450 – Illumina methyl450 array methylation data from the Hudson Alpha Institute for Biotechnology, SYDH – Stanford/Yale/Davis/Harvard, UW – University of Washington, UNC FAIRE – formaldehyde-assisted isolation of regulatory elements from the University of North Carolina, HS – hypersensitivity sites, TFBS – transcription factor binding sites, UTA – University of Texas at Austin, UC – University of Chicago, Txn Factor ChIP – transcription factor chromatin immunoprecipitation.

The JASPAR tool, which contains a repository of matrix-based TFBS, was used to investigate whether rs409558 might be expected to disrupt any TFBS. Two matrices for the sequence flanking rs409558 were analysed, one harbouring the ancestral allele A and one harbouring the minor allele G. For the E2F1 transcription factor, matrix scores predicted a higher binding affinity to the sequence with the G allele (77.3%) than to the sequence with the A allele (64.8%). The sequence containing the G allele forms a near-perfect E2F1 consensus sequence, with only one mismatch at the last base (Figure 3-6). Analysis of ENCODE data for the rs4711279 SNP within the *C6orf48* promoter showed several transcription factor binding sites detected by ChIP-seq, including E2F1 (HeLa-S3 and MCF-7 cells, score of 1000), E2F4 (HeLa-S3 and K562 cells, score of 1000), E2F6 (K562 cells, score of 1000), and MAX (K562 cells, score of 886). Analysis of the rs4711279 flanking sequence using the JASPAR tool did not reveal disrupted TFBS. Nevertheless, these results support a direct regulatory role for rs409558 at *MSH5* and rs4711279 at *C6orf48*, which may explain the *cis*-associations reported for the SLE-protective *MSH5* haplotype in section 3.2.1.

A. **B.**



5′ AAAGC[G]CT 3′
3′ TTTCG[C]GA 5′
1 2 3 4 5 6 7 8

Figure 3-6 E2F1 consensus binding sequence.

A. Sequence logo for the E2F1 consensus binding site showing the requirement for a C/G allele at position 6 (image exported from JASPAR). **B.** Double stranded rs409558 flanking sequence with base position shown at the bottom. Presence of the minor allele G in position 6 (shown in square brackets) forms a near-perfect E2F1 binding site. The mismatched base with the E2F1 consensus binding site at position 8 is highlighted in red.

3.3 *HLA-DRB1* risk haplotypes

3.3.1 *Cis*-eQTLs arising from the *HLA-DRB1*03:01* risk haplotype

To test association between the *DRB1*03:01* haplotype and gene expression, the rs2187668 SNP (located in intron 1 of *HLA-DQA1*) was used as a proxy, since it is in perfect LD with the *DRB1*03:01* allele in Europeans (r^2 = 1.00, U.K. cohort from Fernando et al., 2012).

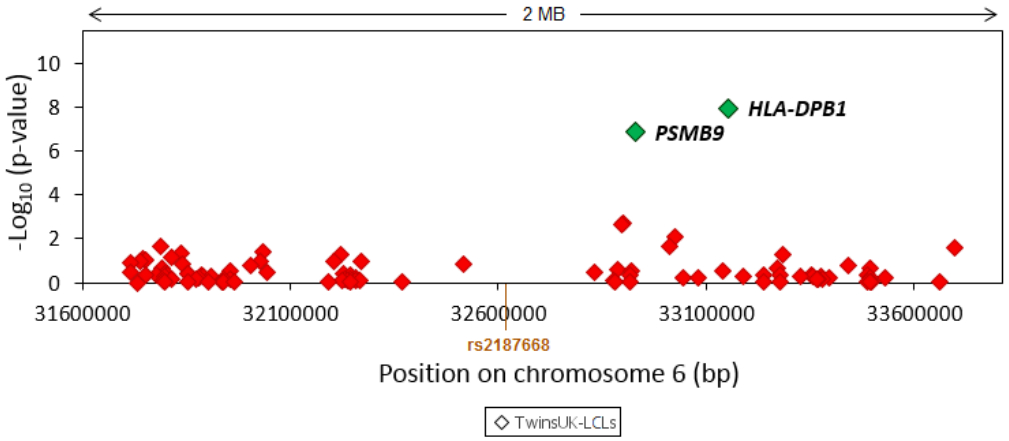


Figure 3-7 Regional plot of *cis*-eQTL associations for the *HLA-DRB1*03:01* tag SNP, rs2187668, in LCLs from the MuTHER cohort (Grundberg et al., 2012).

The strength of association with expression probes (lozenges) located within 1 MB of the SNP is plotted as -$\log_{10}$(p-value) against genomic location. *Cis*-associations significant at an FDR of 5% are coloured in green.

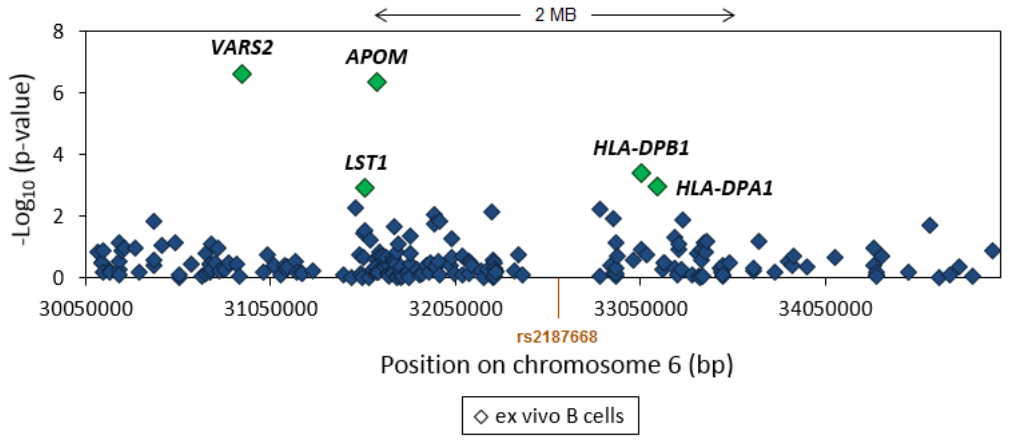


Figure 3-8 Regional plot of *cis*-eQTL associations for the *HLA-DRB1*03:01* tag SNP, rs2187668, in *ex vivo* B cells from Fairfax et al. (2012).

The strength of association with expression probes (lozenges) located within 2.5 MB of the SNP is plotted as -$\log_{10}$(p-value) against genomic location. The arrows indicate the region within 1 MB of the SNP. *Cis*-associations significant at an FDR of 5% are coloured in green.

Cis-eQTL analysis for rs2187668 in MuTHER LCLs showed a significant association with up-regulation of *DPB1* ($P = 1.20 \times 10^{-8}$, FC = 1.17) and *PSMB9* ($P = 1.39 \times 10^{-7}$, FC = 1.09, Figure 3-7). In contrast, no significant eQTLs were found in HapMap LCLs. *Ex vivo* B cells showed several significant *cis*-genes for rs2187668, including *VARS2* ($P = 2.58 \times 10^{-7}$), *APOM* ($P = 4.40 \times 10^{-7}$), *DPA1* ($P = 1.06 \times 10^{-3}$), and *DPB1* ($P = 3.84 \times 10^{-4}$, FC = 1.15, Figure 3-8). Association with *PSMB9* did not reach significance in *ex vivo* B cells following multiple test correction ($P = 5.73 \times 10^{-3}$, FC = 1.10, FDR > 5%); however, a search for published *cis*-eQTLs using HaploReg revealed significant differential expression of *PSMB9* in two monocyte data sets ($P = 2.13 \times 10^{-4}$, Fairfax et al., 2014, and $P = 1.87 \times 10^{-6}$, Zeller et al., 2010) and two whole blood data sets ($P = 1.34 \times 10^{-10}$, Westra et al., 2013, and $P = 3.10 \times 10^{-6}$, Fehrmann et al., 2011).

The *cis*-association of rs2187768 with *DPB1* expression observed in MuTHER LCLs and *ex vivo* B cells was also validated in monocytes ($P = 1.20 \times 10^{-6}$, Fairfax et al., 2014, and $P = 2.10 \times 10^{-15}$, Zeller et al., 2010). The correlation of rs2187668 genotypes with *DPB1* in *ex vivo* B cells showed a heterozygote effect, rather than an additive effect (Figure 3-9). However, this observation is confounded by the small number of rs2187668 homozygotes in this data set (N = 3, Appendix A). As the *DPB1* gene is highly polymorphic, the observed *cis*-eQTL could be a false positive result caused by off-target probe binding. To exclude this possibility, the target sequence for the *cis*-associated *DPB1* Illumina probe, ILMN_1749070, was aligned against the human genome sequence, using the BLAT tool from the UCSC Genome Browser. BLAT analysis showed that the ILMN_1749070 probe did not map to SNP-containing regions in any of the sequenced MHC haplotypes. Figure 3-10 illustrates the probe target site within the *DPB1* gene in the reference MHC haplotype. Table 3-3 summarises all rs2187668 *cis*-genes that were replicated, or showed suggestive evidence of replication, in published data sets from LCLs, *ex vivo* B cells, *ex vivo* monocytes, and whole blood.

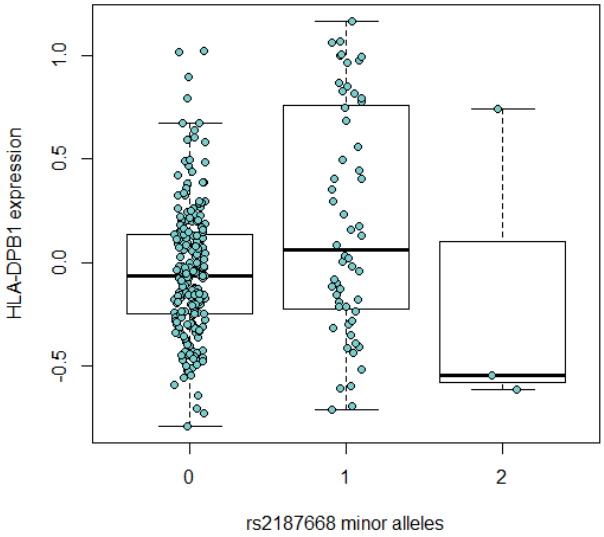


Figure 3-9 eQTL box plot for the *HLA-DRB1*03:01* haplotype tag SNP, rs2187668, and *HLA-DPB1* expression in *ex vivo* B cells from Fairfax et al. (2012).

rs2187668 is a significant *cis*-eQTL for *HLA-DPB1* ($P = 3.84 \times 10^{-4}$, FDR < 5%). The box plot shows median, upper, and lower quartiles of expression levels for each genotype group. Whiskers show the maximal and minimal non-outlying data points.

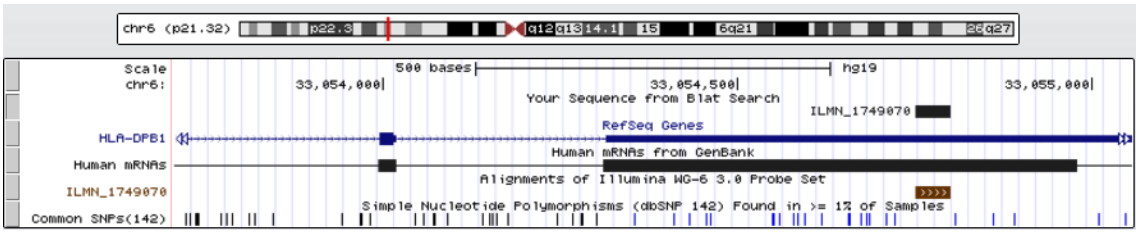


Figure 3-10 UCSC Genome Browser screenshot showing the ILMN_1749070 probe target site at the 3' untranslated region of *HLA-DPB1* in the reference MHC sequence.

The annotation sets shown here include the BLAT search result for ILMN_1749070, the RefSeq gene annotation, the GenBank mRNA annotation, Illumina Sentrix Human-6 BeadChip array probes, and common SNPs from dbSNP (minor allele frequency > 1%).

Table 3-3 Summary of *cis*-eQTLs for rs2187668 (*DRB1*03:01* tag) that were replicated in independent data sets from *ex vivo* B cells, LCLs, *ex vivo* monocytes, and whole blood.

Correlated transcript	***Ex vivo* B cells (Fairfax et al., 2012)**	**MuTHER LCLs (Grundberg et al., 2012)§**	***Ex vivo* naïve monocytes (Fairfax et al., 2014)***	***Ex vivo* monocytes (Zeller et al., 2010)***	**Whole blood (Westra et al., 2013)***	**Whole blood (Fehrmann et al., 2011)***
	N = 281	N = 777	N = 421	N = 1,490	N = 5,311	N = 1,469
HLA-DPB1	✓ P = 3.84×10^{-4} FC = 1.15	✓ P = 1.20×10^{-8} FC = 1.17	✓ P = 1.20×10^{-6} ND	✓ P = 2.10×10^{-15} ND	ND	ND
PSMB9	*P = 5.73×10^{-3}* *FC = 1.10*	✓ P = 1.39×10^{-7} FC = 1.09	✓ P = 2.13×10^{-4} ND	✓ P = 1.87×10^{-6} ND	✓ P = 1.34×10^{-10} ND	✓ P = 3.10×10^{-6} ND
LOC642073	✓ P = 4.74×10^{-6} FC = -1.61	ND	✓ P = 3.22×10^{-9} ND	ND	ND	ND
APOM	✓ P = 4.40×10^{-7} FC = -1.11	ND	✓ P = 1.30×10^{-7} ND	✓ P = 9.93×10^{-7} ND	ND	ND
VARS2	✓ P = 2.58×10^{-7} FC = -1.13	ND	✓ P = 3.17×10^{-12} ND	ND	ND	ND

The significance threshold was set to an FDR of 5% in all data sets. Italicized p-values and fold changes indicate suggestive associations (i.e. significant at P < 0.05 prior to correction for multiple testing, but non-significant at an FDR of 5%). All cohorts included healthy individuals of European ancestry. All data sets were generated using expression microarrays, and were corrected for population structure.
* *Cis*-eQTL data extracted from the HaploReg resource.
§ Data sets adjusted for sex.
N – sample size, FC – fold change (effect size per minor allele), ND – no data available.

A *cis*-eQTL analysis for the Spanish SLE-risk SNP within the *MSH5* locus, rs3131379, which is in moderate LD with *DRB1*03:01* (r^2 = 0.62 in U.K. and r^2 = 0.28 in Spanish populations, cohorts from Fernando et al., 2012) was then performed. No *cis*-associations passed the significance threshold in MuTHER LCLs. In *ex vivo* B cells however, *cis*-associations for rs3131379 substantially overlapped with those observed for rs2187668, as would be expected given the LD between these loci (Figure 3-11). *Cis*-association with *DPB1* expression was significant in *ex vivo* B cells (P = 5.12×10^{-6}, FC = 1.21, Figure 3-12), and was validated in two monocyte data sets (P = 1.15×10^{-8}, Fairfax et al., 2014, and P = 5.90×10^{-18}, Zeller et al., 2010). rs3131379 also demonstrated additional *cis*-genes in *ex vivo* B cells and monocytes

compared to rs2187668, including *ZFP57*, *PSMB9*, and *SKIV2L* (Table 3-4, Fairfax et al., 2014, Zeller et al., 2010).

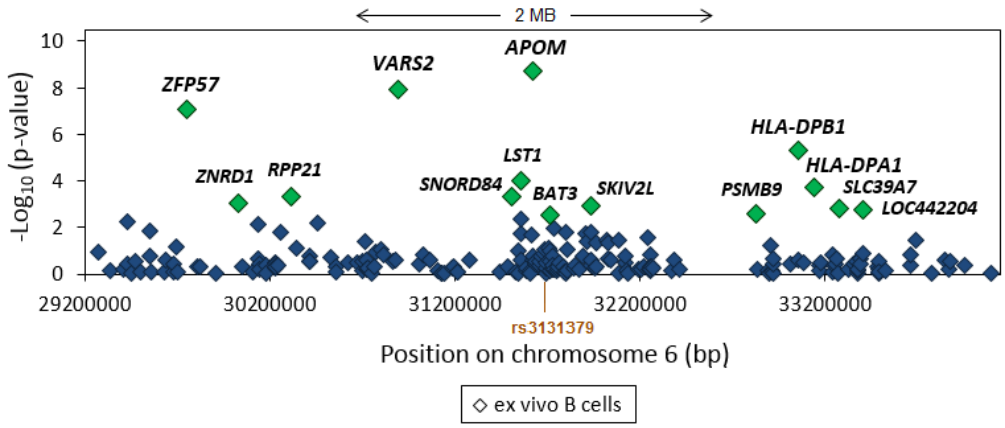


Figure 3-11 Regional plot of *cis*-eQTL associations for the SLE-risk *MSH5* SNP, rs3131379, in *ex vivo* B cells from Fairfax et al. (2012).

The strength of association with expression probes (lozenges) located within 2.5 MB of the SNP is plotted as -log10(p-value) against genomic location. The arrows indicate the region within 1 MB of the SNP. *Cis*-associations significant at an FDR of 5% are coloured in green.

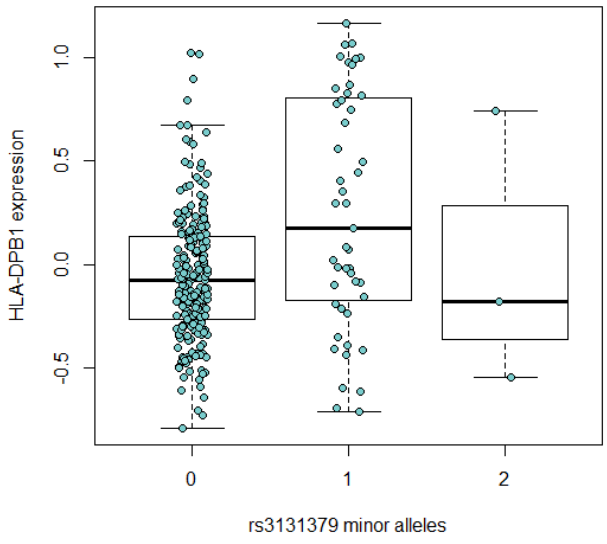


Figure 3-12 eQTL box plot for the *HLA-DRB1*03:01* haplotype tag SNP, rs3131379, and *HLA-DPB1* expression in *ex vivo* B cells from Fairfax et al. (2012).

rs3131379 is a significant *cis*-eQTL for *HLA-DPB1* ($P = 3.84 \times 10^{-4}$, FDR < 5%). The box plot shows median, upper, and lower quartiles of expression levels for each genotype group. Whiskers show the maximal and minimal non-outlying data points.

Table 3-4 Summary of *cis*-eQTLs for rs3131379 that were replicated in independent data sets from *ex* vivo B cells, LCLs, *ex vivo* monocytes, and whole blood.

Correlated transcript	***Ex vivo* B cells (Fairfax et al., 2012)**	***Ex vivo* naïve monocytes (Fairfax et al., 2014)***	***Ex vivo* monocytes (Zeller et al., 2010)***	**Whole blood (Westra et al., 2013)***	**Whole blood (Fehrmann et al., 2011)***
	N = 281	N = 421	N = 1,490	N = 5,311	N = 1,469
HLA-DPB1	✓ 5.12×10^{-6} FC = 1.21	✓ $P = 1.15 \times 10^{-8}$ ND	✓ 5.90×10^{-18} ND	ND	ND
PSMB9	✓ *$P = 5.73 \times 10^{-3}$* *FC = 1.10*	ND	✓ $P = 3.59 \times 10^{-7}$ ND	ND	ND
ZFP57	✓ $P = 4.74 \times 10^{-6}$ FC = -1.61	ND	✓ $P = 1.17 \times 10^{-6}$ ND	ND	ND
SKIV2L	✓ $P = 4.40 \times 10^{-7}$ FC = -1.11	ND	ND	✓ $P = 9.75 \times 10^{-10}$ ND	✓ $P = 2.60 \times 10^{-8}$ ND

The significance threshold was set to an FDR of 5% in all data sets. Italicized p-values and fold changes indicate suggestive associations (i.e. significant at P < 0.05 prior to correction for multiple testing, but non-significant at an FDR of 5%). All cohorts included healthy individuals of European ancestry. All data sets were generated using expression microarrays, and were corrected for population structure.
* *Cis*-eQTL data extracted from the HaploReg resource.
N – sample size, FC – fold change (effect size per minor allele), ND – no data available.

3.3.2 *Cis*-eQTLs arising from the *HLA-DRB1*15:01* risk haplotype

eQTLs for the *DRB1*15:01* risk haplotype were analysed in *ex vivo* cell and LCL data sets using the proxy SNP rs3135391 ($r^2 = 0.98$, U.K. cohort from Fernando et al., 2012), located in exon 2 of *HLA-DRA*. rs3135391 showed significant associations with *BAT2* ($P = 2.81 \times 10^{-5}$) and *PSMB9* ($P = 3.35 \times 10^{-5}$) in MuTHER LCLs (Figure 3-13). In contrast, *ex vivo* B cells showed significant associations with *LOC642073* ($P = 4.33 \times 10^{-22}$) and *ATP6V1G2* ($P = 1.03 \times 10^{-4}$, Figure 3-14), and suggestive association with *BAT2* ($P = 1.32 \times 10^{-2}$, FDR > 5%). The *cis*-eQTLs to *BAT2* and *LOC642073* were validated in naïve monocytes ($P = 4.47 \times 10^{-6}$ and $P = 2.96 \times 10^{-15}$, respectively, Fairfax et al., 2014). None of the other rs3135391 *cis*-associations identified in B cells and LCLs were replicated in publicly-available data sets.

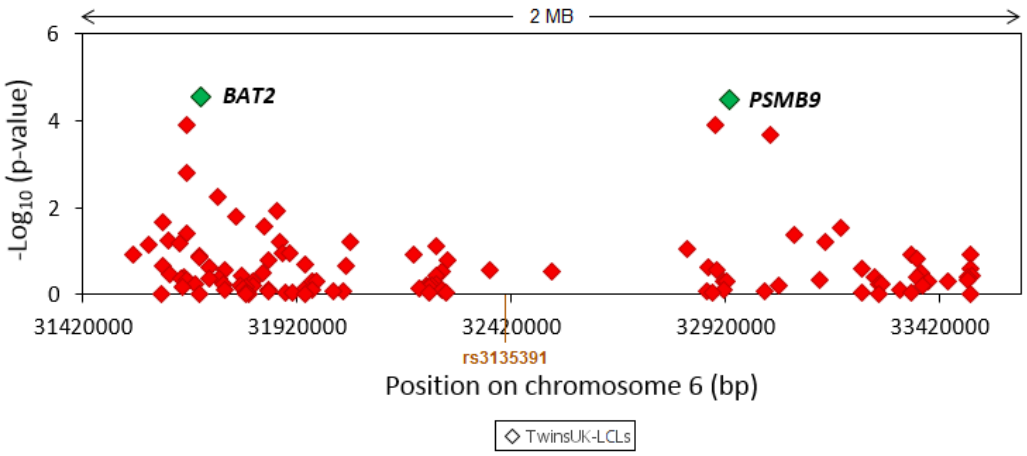


Figure 3-13 Regional plot of *cis*-eQTL associations for the *HLA-DRB1*15:01* tag SNP, rs3135391, in LCLs from the MuTHER cohort (Grundberg et al., 2012).

The strength of association with expression probes (lozenges) located within 1 MB of the SNP is plotted as -$\log_{10}$(p-value) against genomic location. *Cis*-associations significant at an FDR of 5% are coloured in green.

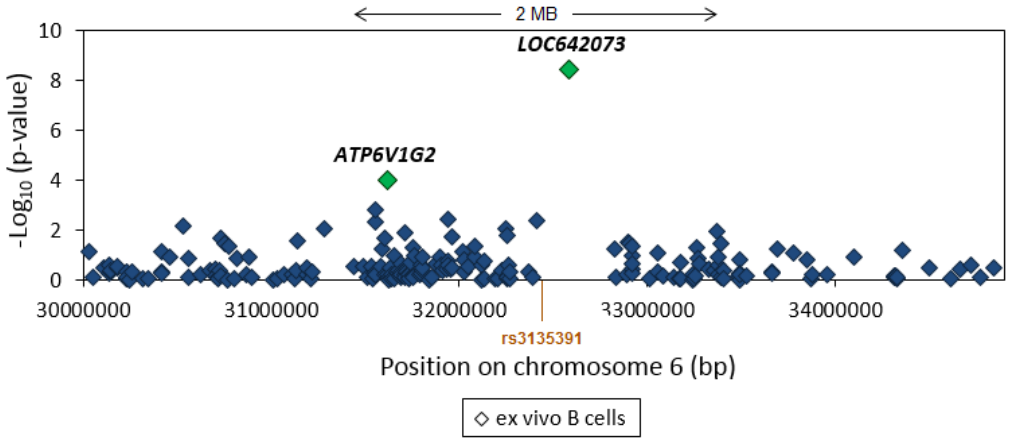


Figure 3-14 Regional plot of *cis*-eQTL associations for the *HLA-DRB1*15:01* tag SNP, rs3135391, in *ex vivo* B cells from Fairfax et al. (2012).

The strength of association with expression probes (lozenges) located within 2.5 MB of the SNP is plotted as -$\log_{10}$(p-value) against genomic location. The arrows indicate the region within 1 MB of the SNP. *Cis*-associations significant at an FDR of 5% are coloured in green.

3.4 *HLA-DPB1* protective and risk haplotypes

3.4.1 LD analysis

The regulatory effects of three *DPB1* variants associated with SLE in different populations (Spanish - rs3117213, Filipino - rs2071351, and African-American - rs2071349) were investigated. In order to assess independence, the LD between the SLE-associated Spanish and Filipino *DPB1* variants and HLA alleles was analysed using genotypic data from Spanish and Filipino cohorts obtained from Fernando and colleagues (Fernando et al., 2012). Both the Spanish and Filipino variants showed low LD with *DRB1* alleles in the respective population ($r^2 < 0.11$). The independence of the African-American *DPB1* variant from HLA alleles could not be established in the African-American population; however, the SNP was in low LD with *DRB1* alleles in the Spanish and Filipino cohorts ($r^2 < 0.14$). Interestingly, all three *DPB1* variants are in low LD with each other, and with *DRB1*03:01* and *DRB1*15:01* haplotype tag SNPs in the U.K. population ($r^2 < 0.19$, cohort from TwinsUK, Figure 3-15).

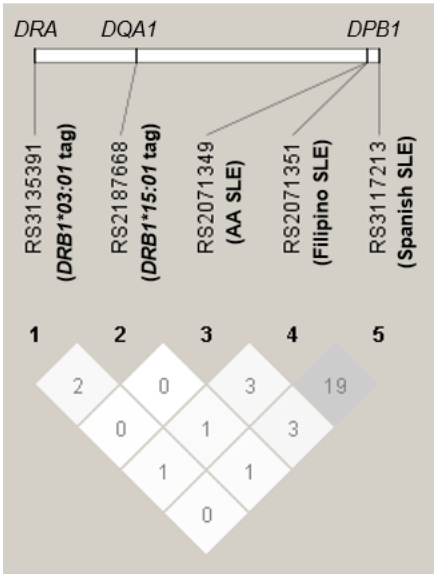


Figure 3-15 LD plot of *HLA-DRB1* and *HLA-DPB1* SLE-associated variants.

The top bar shows the relative location of the European tag SNPs for the *DRB1*03:01* and *DRB1*15:01* SLE-risk haplotypes, the Spanish *DPB1* risk SNP, the Filipino *DPB1* protective SNP, and the African-American (AA) *DPB1* risk SNP. r^2 values from the TwinsUK cohort are displayed below. The LD plot was generated using Haploview version 4.2.

3.4.2 *Cis*-eQTLs arising from *HLA-DPB1* haplotypes

3.4.2.1 The Spanish *HLA-DPB1* haplotype

Genotypes for the *DPB1* SNP rs3117213, which is associated with SLE risk in the Spanish population, correlated significantly with up-regulation of *DPB1* in MuTHER LCLs ($P = 3.08 \times 10^{-49}$) and *ex vivo* B cells ($P = 2.10 \times 10^{-14}$, Figure 3-16). This *cis*-eQTL was validated in LCLs ($P = 1.27 \times 10^{-8}$, GTEx Consortium, 2015), monocytes ($P = 2.09 \times 10^{-24}$, Fairfax et al., 2014), whole blood ($P = 9.81 \times 10^{-198}$, Westra et al., 2013, and $P = 3.18 \times 10^{-27}$, GTEx Consortium, 2015), and in all other tissues tested in the GTEx Consortium study (data not shown, GTEx Consortium, 2015). rs3117213 was also associated at a lesser level of significance with *HLA-DOA* in MuTHER LCLs ($P = 9.35 \times 10^{-7}$), and with the pseudogene *HLA-DPB2* in MuTHER LCLs ($P = 8.74 \times 10^{-5}$), *ex vivo* B cells ($P = 5.73 \times 10^{-5}$), and whole blood ($P = 5.22 \times 10^{-29}$, GTEx Consortium, 2015).

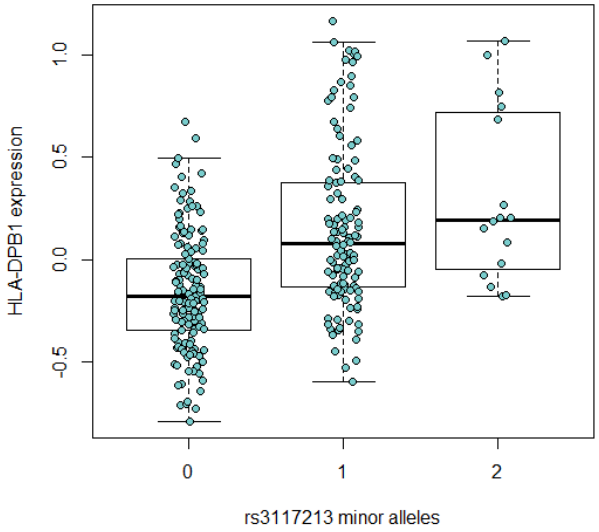


Figure 3-16 eQTL box plot for the Spanish SLE risk-associated *HLA-DPB1* SNP, rs3117213, and *HLA-DPB1* expression in *ex vivo* B cells from Fairfax et al. (2012).

rs3117213 is a significant *cis*-eQTL for *HLA-DPB1* ($P = 2.10 \times 10^{-14}$, FDR < 5%). The box plot shows median, upper, and lower quartiles of expression levels for each genotype group. Whiskers show the maximal and minimal non-outlying data points.

3.4.2.2 The Filipino *HLA-DPB1* haplotype

The association of *DPB1* with SLE protection in Filipinos is conferred by the ancestral allele of rs2071351. Interestingly, *cis*-eQTL analysis for this SNP also showed strong correlation with *DPB1* expression. The SLE risk-associated minor allele correlated significantly with up-regulation of *DPB1* in MuTHER LCLs ($P = 1.47 \times 10^{-49}$), *ex vivo* B cells ($P = 1.63 \times 10^{-22}$, Figure 3-17), monocytes ($P = 6.16 \times 10^{-29}$, Fairfax et al., 2014), and whole blood ($P = 9.81 \times 10^{-198}$, Westra et al., 2013, and $P = 4.65 \times 10^{-11}$, GTEx Consortium, 2015), as well as in all other tissues tested in the GTEx Consortium study (data not shown, GTEx Consortium, 2015). In addition to *DPB1*, rs2071351 genotypes also correlated with *DPA1* expression in MuTHER LCLs ($P = 6.44 \times 10^{-7}$), *ex vivo* B cells ($P = 3.44 \times 10^{-12}$), and monocytes ($P = 4.73 \times 10^{-10}$, Fairfax et al., 2014).

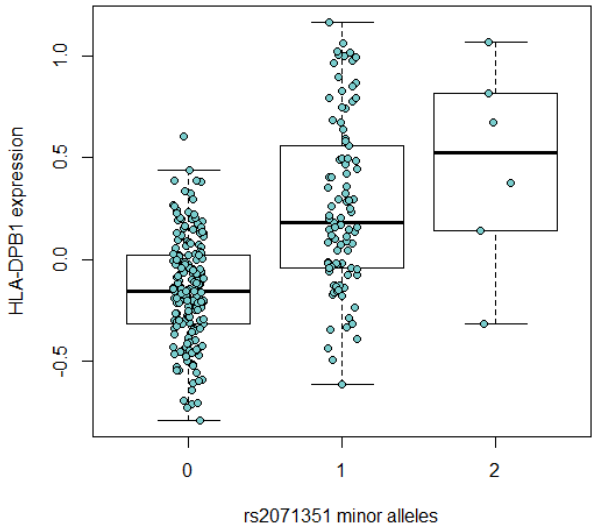


Figure 3-17 eQTL box plot for the *HLA-DPB1* SNP associated with SLE protection in Filipinos, rs2071351, and *HLA-DPB1* expression in *ex vivo* B cells from Fairfax et al. (2012).

rs2071351 is a significant *cis*-eQTL for *HLA-DPB1* ($P = 1.63 \times 10^{-22}$, FDR < 5%). The box plot shows median, upper, and lower quartiles of expression levels for each genotype group. Whiskers show the maximal and minimal non-outlying data points.

3.4.2.3 The African-American *HLA-DPB1* haplotype

In contrast with rs3117213 and rs2071351, the *DPB1* SNP associated with SLE risk in African-Americans, rs2071349, was moderately associated with down-regulation of *DPB1* in MuTHER LCLs ($P = 2.87 \times 10^{-6}$). In *ex vivo* B cells however, association with *DPB1* was not significant after multiple test correction ($P = 6.66 \times 10^{-3}$, FDR > 5%, Figure 3-18). rs2071349 was also associated with expression of the pseudogene *DPB2* in MuTHER LCLs ($P = 1.01 \times 10^{-7}$). Table 3-5 summarises all *cis*-associations with *DPB1* expression that were identified in MuTHER LCLs and *ex vivo* B cells arising from SLE-risk and -protective haplotypes.

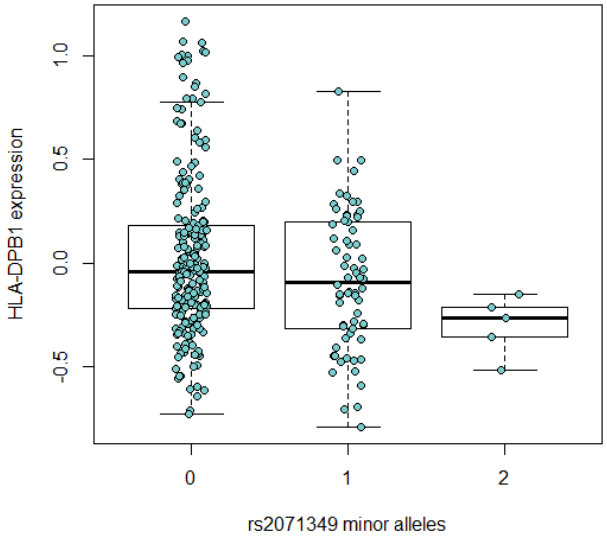


Figure 3-18 eQTL box plot for the African-American SLE risk-associated *HLA-DPB1* SNP, rs2071349, and *HLA-DPB1* expression in *ex vivo* B cells from Fairfax et al. (2012).

rs2071349 is not a significant *cis*-eQTL for *HLA-DPB1* ($P = 6.66 \times 10^{-3}$, FDR > 5%). The box plot shows median, upper, and lower quartiles of expression levels for each genotype group. Whiskers show the maximal and minimal non-outlying data points.

Table 3-5 Summary of *cis*-associations with *DPB1* expression levels for the SLE-associated *DRB1*03:01* haplotype (tagged by rs2187668), and the *DPB1* haplotypes tagged by rs3117213, rs2071351 and rs2071349 in MuTHER LCLs (Grundberg et al., 2012) and *ex vivo* B cells from Fairfax et al. (2012).

SLE-associated SNP (gene)	Associated population	Correlated *cis*-transcript	Cell type	P-value	Fold change	MAF
rs2187668 (*DQA1*)	U.K.	*HLA-DPB1*	LCLs	1.20×10^{-8}	1.17	13.50%
			ex vivo B cells	3.84×10^{-4}	1.15	9.66%
rs3117213 (*DPB1*)	Spanish	*HLA-DPB1*	LCLs	3.08×10^{-49}	1.36	27.02%
			ex vivo B cells	2.10×10^{-14}	1.24	27.04%
rs2071351 (*DPB1*)	Filipino	*HLA-DPB1*	LCLs	1.47×10^{-49}	1.43	17.46%
			ex vivo B cells	1.63×10^{-22}	1.36	19.45%
rs2071349 (*DPB1*)	African-American	*HLA-DPB1*	LCLs	2.87×10^{-6}	-1.14	12.27%
			ex vivo B cells	*6.66×10^{-3}*	*-1.10*	13.52%

The significance threshold was set to an FDR of 5% in all data sets. Italicized p-values and fold changes indicate non-significant p-values.
MAF – minor allele frequency.

3.5 *Trans*-eQTLs within risk and protective MHC haplotypes

Trans-eQTL analysis of *DRB1*, *DPB1*, and *MSH5* SLE haplotypes showed few and inconsistent associations with gene expression in different cell types. In MuTHER LCLs, the *DRB1*03:01* tag SNP showed a significant *trans*-effect to the pseudogene *RNF5P1*, located on chromosome 8 ($P = 1.00 \times 10^{-9}$, FC = 1.10). However, the peak expression SNP (eSNP) for this gene was the MHC class II SNP rs3132965 ($P = 7.61 \times 10^{-20}$, FC = 1.17), which is in moderate LD with the *DRB1*03:01* haplotype ($r^2 = 0.31$, TwinsUK cohort). In contrast, no significant *trans*-eQTLs for the *DRB1*03:01* haplotype were detected in *ex vivo* B cells.

The *DRB1*15:01* tag SNP showed a significant *trans*-association with *ERG*, located in chromosome 21, in MuTHER LCLs ($P = 5.98 \times 10^{-23}$, FC = -1.10). The peak *trans*-signal for *ERG* was to the class II SNP rs7194 ($P = 5.41 \times 10^{-50}$, FC = -1.12). This SNP is in low to moderate LD with the *DRB1*15:01* haplotype ($r^2 = 0.25$, TwinsUK cohort). The *DRB1*03:01* and *DRB1*15:01* *trans*-associations with *RNF5P1* and *ERG* expression may therefore result from LD with the causal variants rs3132965 and rs7194, respectively. *Trans*-eQTL analyses in *ex vivo* B cells showed two significant associations for the *DRB1*15:01* tag SNP, *ZNF672* in

chromosome 1 (P = 1.93 × 10^{-6}, FC =-1.06) and *FAM98B* in chromosome 15 (P = 2.25 × 10^{-6}, FC = 1.12). Interestingly, the *trans*-association to *ZNF672* was replicated in naïve monocytes (P = 2.11 × 10^{-22}, Fairfax et al., 2014). The peak *trans*-signal for *ZNF672* in this monocyte data set was to the class II SNP rs3104369 (P = 3.75 × 10^{-31}), which is in moderate LD with the *DRB1*15:01* haplotype (r^2 = 0.23, TwinsUK cohort). None of the tag SNPs for *DRB1*03:01*, *DRB1*15:01*, *DPB1* (rs3117213, rs2071351, rs2071349), or *MSH5* (rs409558, rs3131379) haplotypes showed any other significant *trans*-associations in the studied data sets.

3.6 Discussion

This chapter describes a comprehensive investigation of *cis*- and *trans*-eQTLs within SLE-associated MHC haplotypes in LCLs and *ex vivo* B cells, using publicly-available data sets. The *DRB1*03:01* risk haplotype and the *MSH5* protective haplotype in class III were significant *cis*-eQTLs for multiple transcripts, which were validated in monocytes and whole blood. The *DRB1*03:01* haplotype, and two *DPB1* risk and protective SNPs, showed robust *cis*-associations with the classical HLA class II gene *DPB1* in LCLs, *ex vivo* B cells, and whole blood. In contrast, *cis*-eQTLs arising from the *DRB1*15:01* haplotype, and *trans*-eQTLs arising from all analysed MHC haplotypes, were inconsistent between the different data sets.

Genotypes of the SLE-protective *MSH5* SNP, rs409558, correlated significantly with expression levels of the class III genes *DOM3Z* and *C6orf48* in LCLs, *ex vivo* B cells, monocytes, and whole blood, suggesting an unprecedented role for these genes in SLE. The functions of *DOM3Z* (also known as decapping exoribonuclease, or *DX0*) and *C6orf48* are currently unknown (Entrez Gene). *C6orf48* is moderately expressed in human B cells and highly expressed in lymphoma tumours, whereas *DOM3Z* is thought to be a housekeeping gene (Pan et al., 2012, BioGPS, Entrez Gene). Furthermore, the *cis*-association with *C6orf48* was replicated in skin samples from two independent cohorts, which is of interest given the cutaneous clinical manifestations sometimes observed in lupus patients. In *ex vivo* B cells, monocytes, and whole blood, rs409558 was also a significant *cis*-eQTL for the class III gene *HSPA1B*, which has previously shown association with SLE risk independently of *DRB1*03:01* in Europeans and African Americans (Furnrohr et al., 2010, Jarjour et al., 1996). The suggestive

cis-eQTL to the *MSH5* gene itself in B cells and LCLs is of interest, given the association of variants near this gene with autoimmune diseases, and the role of *MSH5* in pathways relevant to SLE (such as deoxyribonucleic acid (DNA) repair, Bannwarth et al., 2012). However, the *cis*-association to *MSH5* was only suggestive for two of five available *MSH5* probes. The lack of power to detect very small fold changes, as well as the presence of potential spliced transcripts not targeted by the Illumina array, may explain the lack of significance for these *cis*-associations.

The SLE-protective *MSH5* haplotype was defined by thirteen SNPs, and was shown to be independent of HLA alleles in European populations. Bioinformatic analysis of the thirteen SNPs delineating the haplotype indicated that both rs409558 at *MSH5* and rs4711279 at *C6orf48* were the best functional SNP candidates within the haplotype. The overlap of rs409558 and rs4711279 with high numbers of regulatory features, including elements indicative of active promoters, may explain the *cis*-eQTLs observed for the *MSH5* protective haplotype, particularly to *C6orf48*. The rs409558 minor allele (G) was predicted to increase binding affinity of E2F1, suggesting that this transcription factor may be functionally involved in regulating gene expression within the protective haplotype. Altogether, the *cis*-eQTLs and functional annotations here presented support a direct regulatory role for non-coding SNPs within the protective *MSH5* haplotype, which could underlie its association with SLE. Following these analyses, a large SLE GWAS was conducted in a cohort of Northern and Southern European ancestry (Bentham et al., 2015). The independent association of rs409558 with SLE could not be validated in this study, due to the SNP failing quality control (Dr David Morris, King's College London; personal communication). Provided that the rs409558 association with SLE protection is validated in independent data sets, the regulatory role of this haplotype could be further investigated using functional studies. For example, differential binding of E2F1 to rs409558 allelic variants may be validated by electrophoretic mobility shift assay (EMSA), and the regulatory role of variants in weak LD with rs409558 could be assessed.

Whereas the *cis*-eQTLs for the *MSH5* protective haplotype were confined to the class III region, the *cis*-eQTL landscape for the ancestral *DRB1*03:01* haplotype spanned both class III and class II regions in *ex vivo* B cells, reflecting the large size of the haplotype. The *DRB1*03:01* haplotype demonstrated reproducible *cis*-eQTL for *DPB1* in LCLs, *ex vivo* B cells, and

monocytes. Furthermore, this *cis*-eQTL showed a possible heterozygote effect in *ex vivo* B cells. This may have been an artefact resulting from the low number of *DRB1*03:01* homozygote samples in the *ex vivo* B cell data set (Appendix A); however, it was not possible to access genotypic data from additional data sets to confirm this. Up-regulation of *DPB1* in *DRB1*03:01* haplotypes is of great interest, given the independent association of the *DPB1* gene with SLE, as well as other autoimmune diseases, in different populations. Interestingly, Spanish and Filipino SLE-associated SNPs in *DPB1*, which are independent of the *DRB1*03:01* haplotype, were also strongly associated with *DPB1* expression in *ex vivo* B cells (up-regulation), LCLs, and whole blood. Furthermore, the *cis*-eQTL within the *DPB1* and *DRB1*03:01* haplotypes showed similar fold changes for differential expression of *DPB1* (FC ~ 1.3, Table 3-5). Since both the Spanish and Filipino *DPB1* SNPs were shown to be in low LD with HLA alleles (including *DRB1*03:01*), the *cis*-associations of SLE-risk haplotypes with increased *DPB1* expression suggest an important role for this class II gene in susceptibility to autoimmune disease. One can hypothesise a plausible pathogenic mechanism for these disease-associated *cis*-eQTLs that would decrease tolerance to self: up-regulation of *DPB1* expression in autoreactive B cells harbouring *DRB1*03:01* or *DPB1* risk haplotypes would result in increased MHC class II presentation of self-antigens to T cells. In the presence of co-stimulatory signals, T cell recognition of self-antigen-MHC class II complexes induces plasma cell differentiation and, subsequently, autoantibody production. Further fine-mapping of the SLE-associated Spanish and Filipino *DPB1* signals is required in order to find the causal regulatory variants.

The *DRB1*03:01* haplotype was also a significant *cis*-eQTL for *PSMB9* in LCLs, monocytes, and whole blood, at similar fold changes. Interestingly, this gene showed additional *cis*-associations with the SLE-risk *MSH5* SNP, rs3131379, which is in moderate LD with *DRB1*03:01*, and with the *DRB1*15:01* haplotype. *PSMB9* is a type 1 interferon (IFN)-inducible gene which encodes a proteasome subunit involved in antigen processing. Further investigation of *PSMB9 cis*-associations with SLE-risk *DRB1* and *MSH5* haplotypes may be more relevant in the context of type 1 IFN stimulation, given the role of *PSMB9* and the importance of IFN-α in the pathogenesis of lupus. In *ex vivo* B cells and monocytes, the *MSH5* risk SNP, rs3131379, showed additional *cis*-associations with *SKIV2L* in class III, which has previously shown an independent association with SLE (Fernando et al., 2007), and to *ZFP57* in class I. Increased

levels of *ZFP57* expression have been previously reported in *B*08-DRB1*03:01*-homozygous LCLs (COX), compared to other MHC-homozygous LCLs (including *B*18-DRB1*03:01*, Vandiedonck et al., 2011). This suggests that the rs3131379 and *B*08-DRB1*03:01 cis*-eQTLs to *ZFP57* are tagging an effect from causal variants in the class I region, which has been shown to harbour regulatory *cis*-SNPs in *ZFP57* (Plant et al., 2014).

Trans-eQTLs within the *DRB1*03:01* and *DRB1*15:01* haplotypes showed significant but inconsistent associations with gene expression in *ex vivo* B cells and LCLs. These associations may result from weak LD between the peak eSNPs for the *trans*-associated genes and *DRB1* haplotypes. With the exception of the *trans*-association of *DRB1*15:01* to *ZNF672*, none of the *trans*-signals were replicated in other data sets, rendering these results inconclusive. Replication of *trans*-signals in different cell types is difficult due to the high context-specificity of *trans*-eQTLs. This was exemplified in an eQTL study by Fairfax and colleagues showing low overlap of *trans*-eQTLs in *ex vivo* B cells and monocytes (N = 120), compared to *cis*-eQTLs (n = 17,962, Fairfax et al., 2012).

This study has highlighted some of the caveats of conducting eQTL analyses using expression arrays, particularly with respect to the MHC region. Firstly, the sample size in the data sets investigated in this study may have been too low to enable detection and replication of *cis*- and *trans*-associations of small effect size (particularly for *DRB1*03:01* and *DRB1*15:01* haplotypes, for which the number of homozygotes was low). Failure to replicate *cis*- and *trans*-associations may also be due to the differences in cell type and activation states. The *ex vivo* B cell data set was derived from positively selected B cells, so it is possible that the cells may have been partially activated during the selection process. LCLs are immortalized via EBV-transformation, which also results in cell activation. As eQTLs are highly sensitive to the cellular context, comparing eQTLs between the B cell, LCL, monocyte, and whole blood data sets will fail to replicate cell state- and cell type-specific effects. Furthermore, most of the eQTL data sets investigated in this study were not adjusted for sex, which is another factor known to influence eQTLs. Another limitation to the study of gene expression at the MHC and genome-wide is the use of Illumina expression arrays for the generation of the LCL and *ex vivo* cell data sets. Current Illumina arrays are designed to target the 3' untranslated region of genes and a limited number of annotated spliced transcripts, therefore any *cis*- and *trans*-effects on spliced

transcript variants may be missed. Detection of gene expression at the MHC by microarray or RNA sequencing methods is also limited by the high polymorphism and sequence homology in the region. As conventional microarrays are designed to target the reference human genome MHC sequence (which harbours the *B*07:02-DRB1*15:01* haplotype), this results in the exclusion of several HLA genes, as well as copy number variable genes, from most RNA sequencing and non-custom array studies.

One other important caveat in this study is that the SLE-associated signals in *MSH5* and *DPB1* here investigated may be erroneous, due to the relatively small size of the cohorts in which they were identified. Therefore, these associations require validation in independent cohorts of the relevant ancestry, prior to any further functional investigations. Furthermore, the use of European gene expression data to annotate eQTLs for SNPs associated with SLE in non-European populations (such as Filipino and African-American populations) may result in inaccurate eQTL signals, due to differences in eQTL landscape between populations. Importantly, the SNPs here investigated may not be the peak eSNPs for the genes where a probe association was found. In order to identify the most likely causal variants for the eQTL signals identified in this chapter, the peak *cis*-eQTL for each *cis*-transcript here reported, as well as their LD with the SLE-associated SNPs under investigation, should also be explored.

In conclusion, this study demonstrates eQTLs within SLE-associated *DRB1*, *DPB1*, and *MSH5* haplotypes in LCLs and B cells, which were reproducible in different blood cell types. These eQTLs suggest that SLE-risk and -protective haplotypes affect disease risk through gene regulatory mechanisms, and highlight new candidate genes in SLE. Thus, the regulatory role of disease-associated variants at the MHC here presented supports the hypothesis that disease associations arising from the MHC region, particularly *DRB1* haplotypes, result from the combinatorial effect of HLA alleles and regulatory variants.

Chapter 4 - The effect of IFN-α on the B cell transcriptome

4.1 Introduction

It is well established that type 1 interferons (IFNs), particularly IFN-α, are central to the pathogenesis of systemic lupus erythematosus (SLE). Many of the immunological and clinical features of the disease are attributed to an autoimmune response driven by type 1 IFNs. As mentioned in Chapter 1, multiple lines of evidence support a role for type 1 IFNs in triggering, amplifying, and maintaining inflammation in lupus and in other autoimmune diseases, but the underlying mechanisms remain unclear. Over the past few years, genetic studies have identified signalling components of the type 1 IFN pathway and regulators of nucleic acid metabolism as key players in the pathogenesis of lupus and other systemic autoimmune diseases (Crow, 2010).

B cell dysfunction is also critically implicated in the pathophysiology of SLE. B cell functions thought to be important in SLE have been postulated to include self-antigen presentation to CD4+ T cells, production of pathogenic auto-antibodies, and secretion of pro-inflammatory mediators. Given that IFN-α is a trigger of autoimmunity, it is likely to contribute significantly to B cell-mediated pathology in SLE. Type 1 IFNs are important mediators between the innate and adaptive immune systems, and direct type 1 IFN stimulation of B cells has been shown to be crucial for antibody-mediated response to viral and microbial infections in mice (Coro et al., 2006, Swanson et al., 2010). The effects of IFN-α and IFN-β stimulation on gene and protein expression have been extensively studied in dendritic cells (DCs) and murine B cells (Braun et al., 2002, Hervas-Stubbs et al., 2011), whilst studies in human B cells have looked only at the effects of IFN-α on a limited number of gene transcripts (de Goer de Herve et al., 2011, Giordani et al., 2009). To date, no studies have investigated the impact of IFN-α on global gene expression in human B cells.

Illumina expression microarrays have been a popular choice for transcriptomic data analysis, as exemplified by the study from Fairfax and colleagues, and the Multiple Tissue Human Expression Resource (MuTHER) study, discussed earlier in Chapter 3 (Fairfax et al., 2012, Grundberg et al., 2012). Among the limitations of the Illumina platform design are the targeting

of gene 3' untranslated regions (UTRs), and restricted coverage of annotated splice variants. The Affymetrix Human Exon 1.0 ST array (here referred to as exon array) provides an attractive alternative, due to its high-density gene coverage. The probes are designed to detect individual exons, enabling the quantification of both whole gene expression and alternatively spliced transcripts (Clark et al., 2007). Alternative splicing, like gene expression, is tissue- and context-specific, and can therefore be modulated by pro-inflammatory cytokines (Le et al., 2004).

Investigating the consequences of type 1 IFN stimulation on gene regulation (i.e. gene expression and alternative splicing) in the context of B cells is important to further our understanding of the pathways and mechanisms that are dysregulated in antibody-mediated autoimmune diseases. The work described here seeks to investigate the effect of IFN-α stimulation on the regulation of SLE-associated genes in B cells, in order to elucidate the functional mechanisms underlying these associations.

This chapter describes the first transcriptomic analysis of human B cells following stimulation with IFN-α. First, changes in genome-wide gene expression between resting and IFN-α-stimulated B cells were analysed. Secondly, the differential expression of genes associated with autoimmune and non-autoimmune diseases was investigated, with a particular focus on genes within SLE-associated loci, and within the major histocompatibility complex (MHC) region. Lastly, the impact of IFN-α on alternative splicing in B cells was explored. Given the gender bias of SLE incidence, and the influence of hormonal factors on gene expression, the study cohort comprised 49 healthy women of European ancestry from the TwinsUK registry. Transcriptomic data were generated from negatively selected *ex vivo* CD19+ B cells at rest, and following stimulation with IFN-α for 6 hours. This time point was previously shown to induce the highest expression of four type 1 IFN-inducible genes in a multi-time point assay (work conducted by Ms Lora Boteva, under the supervision of Dr Michelle Fernando, King's College London). Gene expression data from CD19+ B cells were then processed using the Affymetrix exon array. P-values were considered significant if they fell below a false discovery rate (FDR) < 1% or 5%, depending on the analysis.

4.2 Quality control of exon array data

The quality of the expression data obtained from exon arrays was first investigated using the Affymetrix Expression Console software, build 1.2.1.20. Visualization of the quality control (QC) metrics described in Chapter 2, section 2.4.1, revealed that all arrays generated from *ex vivo* B cells had high hybridization quality, and showed normal distribution of probe intensity signals (Figure 4-1).

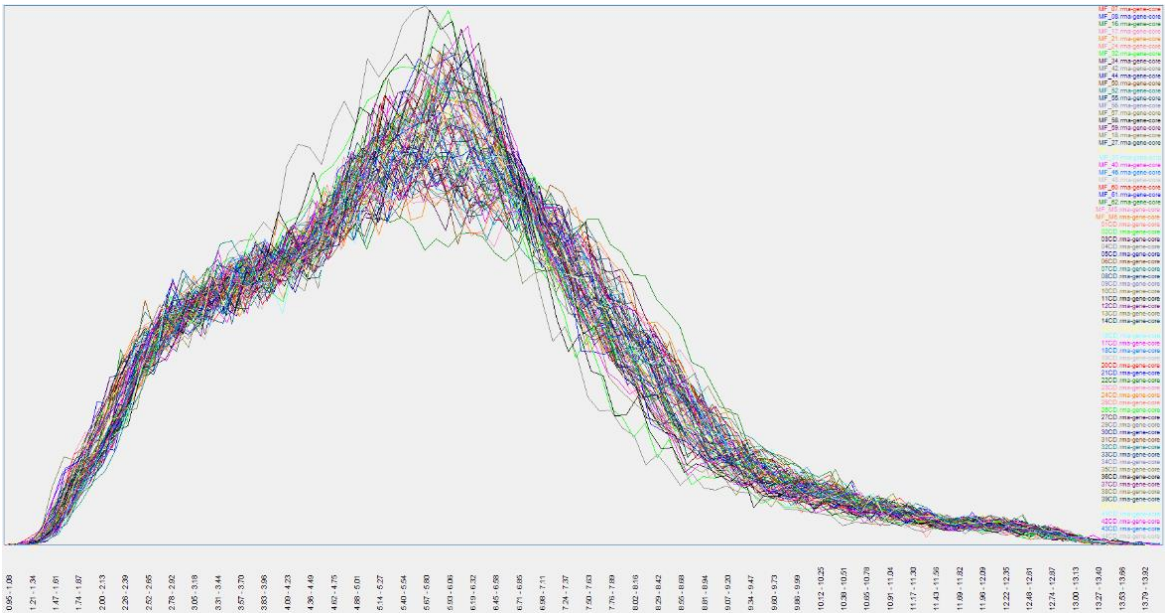


Figure 4-1 Histogram showing the normal distribution of exon array probe set intensity signals for *ex vivo* B cell samples.

Figure generated using the Affymetrix Expression Console software, build 1.2.1.20.

Two technical replicates derived from a single monozygotic twin pair in each experimental batch were available. Within a batch, each twin from the pair was used as a biological replicate. A Spearman correlation test showed high correlation between biological replicates (r^2 = 0.98), and slightly lower correlation between technical replicates (r^2 = 0.89), reflecting a probable batch effect.

Principal component analysis (PCA) of the raw data showed that three principal components (PCs) explained most of the data variability (Figure 4-2). Data structure was investigated using a PCA plot. Samples from different batch years clustered distinctly along PC1 and PC3, and samples from different treatment groups clustered along PC2 (Figure 4-3 A and B). The PCA plot also identified four outlier samples, which were removed from the data prior to analysis. Following adjustment for batch effects, samples from different treatment groups still clustered along PC2 (Figure 4-3 C and D).

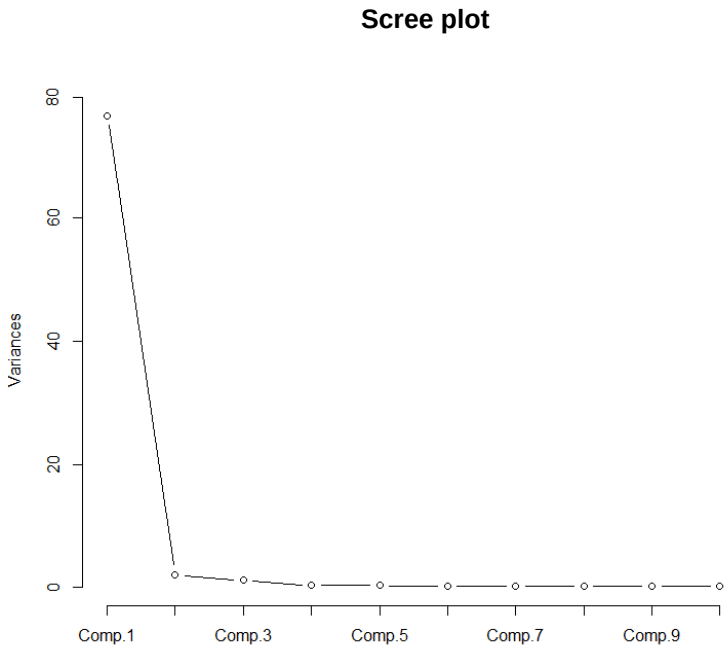


Figure 4-2 Scree plot for exon-level array data from *ex vivo* B cells.

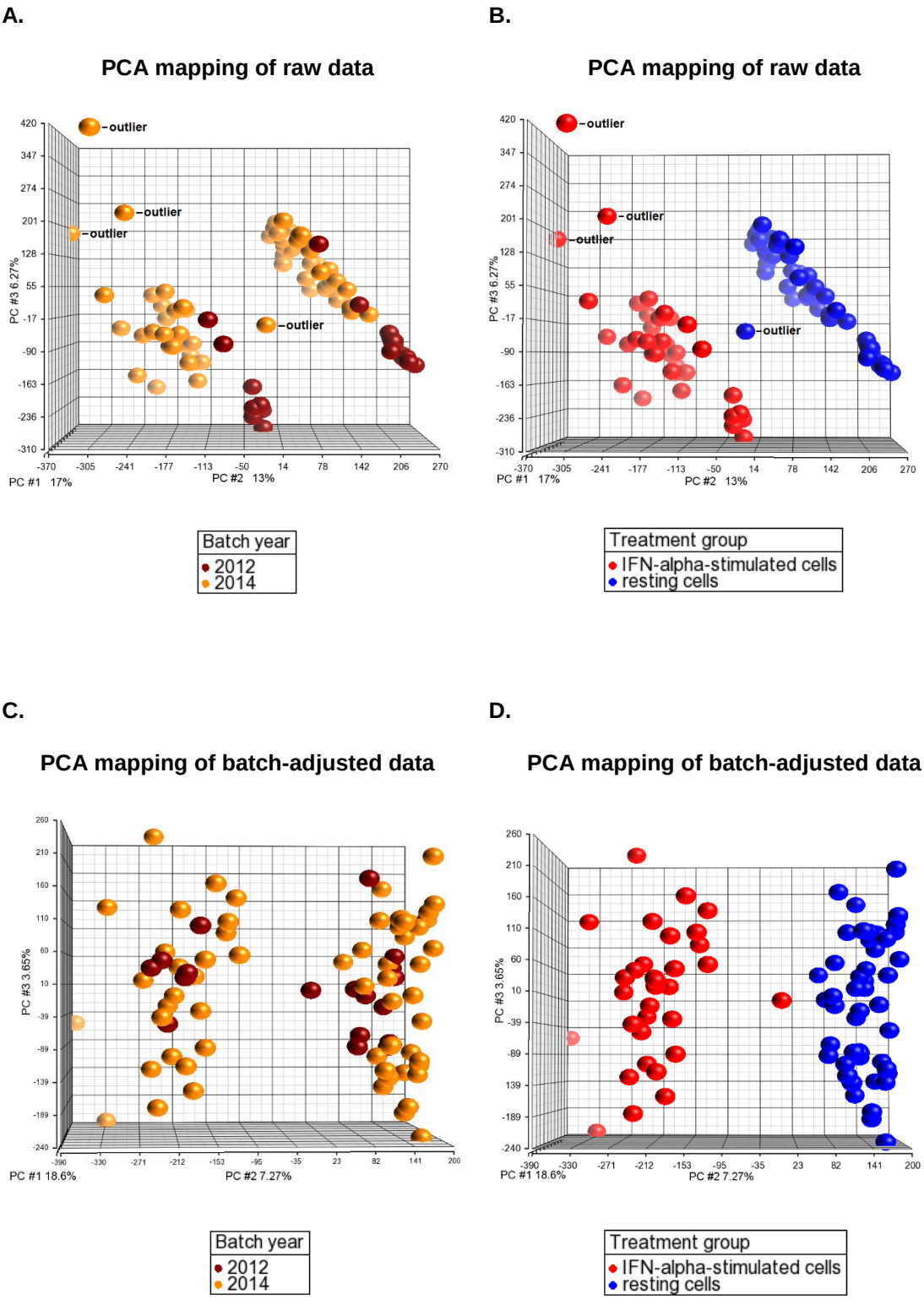


Figure 4-3 PCA plots for exon-level array data from *ex vivo* B cells.

Plot **A** shows samples clustering by batch year in the raw data, and plot **B** shows samples clustering by treatment group in the raw data. Sample outliers are indicated. Plots **C** and **D** show the data structure following adjustment for batch effects and removal of outliers.

4.3 Global effect of IFN-α on gene expression

Gene-level analysis of B cell expression data revealed that IFN-α significantly affected genome-wide gene expression, with 40% of annotated transcripts (6,018 out of 15,510) differentially expressed between resting and IFN-α-stimulated cells (FDR < 1%, Figure 4-4). Of these, 1,126 genes had fold changes ≥ 2 or ≤ -2 (which equates to $\log_2$ fold changes ≥ 1 or ≤ -1). As expected, the most significantly up-regulated genes were IFN response genes, and the most down-regulated genes were primarily involved in cell cycle regulation (e.g. *MOB3B*, *DUSP1*), tumour suppression (e.g. *PLPP5*, *FCRL1*), or innate and adaptive immunity (*LYZ*, *BIN2*, *CD1C*).

Enrichment analysis of the differentially expressed data using Ingenuity Pathway Analysis (IPA) identified B cell receptor (BCR) signalling as the top canonical pathway (56%, $P = 2.13 \times 10^{-12}$), followed by phosphoinositide 3-kinase (PI3K) signalling in B lymphocytes (57%, $P = 1.54 \times 10^{-9}$), which occurs downstream of the type 1 IFN receptor. The IFN signalling pathway was also significantly enriched in cells stimulated with IFN-α (48%, $P = 3.39 \times 10^{-2}$). The top five enriched canonical pathways in the gene set are listed in Table 4-1. The most significant upstream regulators of differential expression predicted by IPA were IFN-λ (*IFNL1*, $P = 3.47 \times 10^{-14}$), IFN-α (*IFNA2*, $P = 2.35 \times 10^{-11}$), and IFN-γ (*IFNG*, $P = 1.43 \times 10^{-9}$).

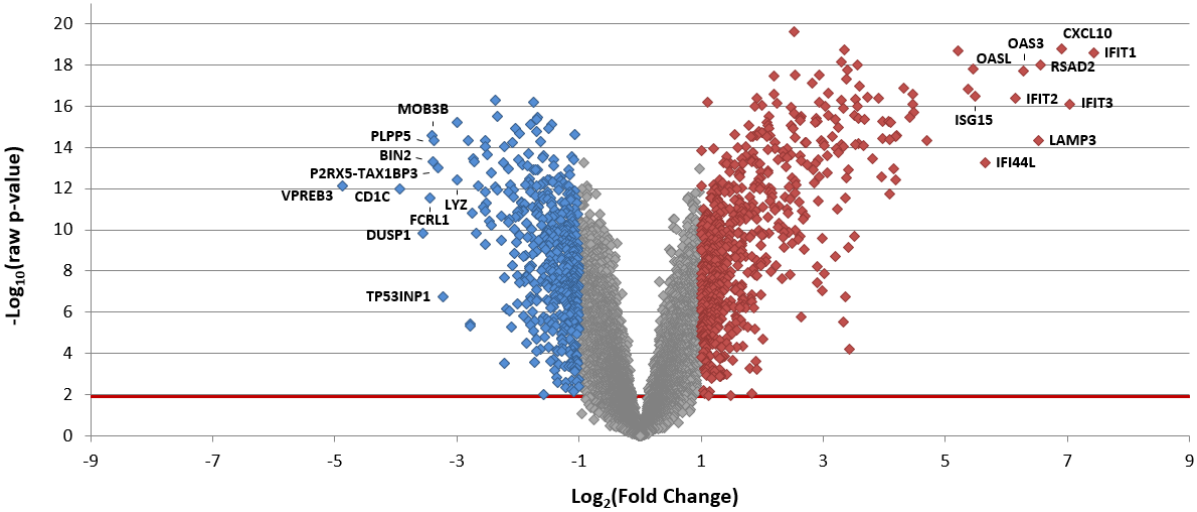


Figure 4-4 Volcano plot showing the effect of IFN-α stimulation on global gene expression in *ex vivo* B cells.

Global gene expression was quantified using the Affymetrix Exon 1.0 ST array. Up-regulated genes with $\log_2$ fold changes ≥ 1 are highlighted in red. Down-regulated genes with $\log_2$ fold changes ≤ -1 are highlighted in blue. Gene labels indicate the top ten up-regulated and down-regulated transcripts. The red line indicates the significance threshold at an FDR of 1%.

Table 4-1 Top five IPA canonical pathways enriched in the gene set differentially expressed by IFN-α.

Canonical Pathway	Number of pathway genes	Number DE transcripts	Percentage DE genes	Z-score	$P_{enrichment}$
B Cell Receptor Signalling	141	79	56.0%	ND	2.13×10^{-12}
PI3K Signalling in B Lymphocytes	96	55	57.3%	-3.159	1.54×10^{-9}
Role of NFAT in Regulation of the Immune Response	116	61	52.6%	-3.015	1.83×10^{-8}
CD28 Signalling in T Helper Cells	84	47	56%	-2.846	6.48×10^{-8}
Molecular Mechanisms of Cancer	297	125	42.1%	ND	9.51×10^{-8}

Gene transcripts differentially expressed by IFN-α (FDR < 1%) were put forward for IPA analysis of canonical pathways. Z-scores denote the predicted direction of change, with significance considered at an absolute z-score of ≥ 2. Pathway enrichment was considered significant if P < 0.05 at an FDR < 5%. DE – differentially expressed, PI3K – phosphoinositide 3-kinase, NFAT – nuclear factor of activated T cells, ND – data not available.

4.3.1 Type 1 IFN signature

The presence of a type 1 IFN signature was detected in IFN-α-stimulated B cells. Five IFN-inducible genes commonly used to characterize the signature, *IFIT1*, *IRF7*, *ISG15*, *IFI44*, and *IFI44L* (Greenberg et al., 2012, Kyogoku et al., 2013, Ferreira et al., 2014), were up-regulated in the exon array data (FDR < 1%). The direction of effect for these genes was validated by quantitative reverse-transcription polymerase chain reaction (qPCR), using complementary deoxyribonucleic acid (cDNA) from the exon array samples (Table 4-2). Interestingly, exon array data showed that down-regulation of the IFN-α receptors 1 (*IFNAR1*) and 2 (*IFNAR2*) messenger ribonucleic acid (mRNA) transcripts on IFN-α stimulation trended towards significance, suggesting a negative feedback loop ($P_{FDR} = 0.074$, fold change (FC) = -1.16 and $P_{FDR} = 2.38 \times 10^{-11}$, FC = -2.52, respectively).

Table 4-2 Up-regulation of five type 1 IFN signature genes after IFN-α stimulation.

Type 1 IFN response gene	Exon array		qPCR	
	Fold change	P_{FDR}	Fold change	P_{FDR}
IFIT1	172.87	2.66×10^{-16}	1254.43	5.63×10^{-18}
IRF7	5.31	5.10×10^{-13}	40.53	2.23×10^{-18}
ISG15	45.01	7.39×10^{-15}	20.80	3.98×10^{-5}
IFI44	26.04	2.48×10^{-13}	13.99	2.35×10^{-5}
IFI44L	50.24	1.90×10^{-12}	12.08	4.65×10^{-5}

Transcripts were considered significantly differentially expressed at an FDR < 1%.

4.3.2 Differential expression of B cell markers by IFN-α

Given that IFN-α strongly affected global gene expression, and that stimulation with cytokines is known to modify the cell phenotype, the biological state of *ex vivo* B cells following stimulation with IFN-α was assessed. B cell development and differentiation occurs in stages, each of which is characterized by different sets of markers expressed at the cell surface. In line with the activation of BCR signalling highlighted above, transcripts of the B cell activation markers *CD25* (*IL2RA*), *CD40*, and *CD80* (Chang et al., 2008, Gantner et al., 2003, Gupta et al., 2013, Breen et al., 2011) were up-regulated by stimulation with IFN-α (Table 4-3). In contrast, other known B cell activation markers, such as the co-stimulatory markers *CD86* and *CD69*, or the memory B cell marker *CD27*, were down-regulated, or not differentially expressed. Interestingly, IFN-α induced differential expression of markers associated with a transition into plasmablasts ($CD20^{-}CD38^{++}CD45^{++}CD126^{++}CD138^{-}$ cells, De Vos et al., 2006), such as down-regulation of *CD20* (*MS4A1*) and *CD22*, and up-regulation of CD38 transcripts (Table 4-4). Similarly, the data showed differential expression of a number of plasma cell markers ($CD20^{-}CD38^{++}CD45^{weak}CD54^{+}CD126^{+}CD130^{+}138^{++}$ cells, De Vos et al., 2006), including down-regulation of *CD45*, and up-regulation of *CD54* (*ICAM1*) and *CD130* (*IL6ST*) transcripts. Differential gene expression did not correlate with B effector cell polarization, as there was no up-regulation of B effector 1 (Be1) or Be2-associated cytokines (de Goer de Herve et al., 2011, Lund and Randall, 2010, Table 4-4). Transcripts were considered significantly differentially expressed at an FDR < 1% for all aforementioned genes.

Table 4-3 The effect of IFN-α stimulation on the expression of B cell activation markers.

B cell activation marker	Fold change	P_{FDR}
CD25 (IL2RA)	**6.11**	**8.00×10^{-11}**
CD40	**1.67**	**4.74×10^{-5}**
CD80	**1.47**	**5.37×10^{-3}**
CD69	1.15	2.12×10^{-1}
CD27	-1.16	1.64×10^{-1}
CD86	-1.13	1.10×10^{-1}

Rows in bold type denote differentially expressed transcripts significant at an FDR < 1%.

Table 4-4 The effect of IFN-α stimulation on the expression of plasmablast, plasma cell, and B effector cell differentiation markers.

B cell differentiation markers	Fold change	P_{FDR}
Plasmablast and plasma cell markers		
***CD130* (*IL6ST*)**	**3.62**	**2.60×10^{-10}**
CD22	**-3.43**	**1.02×10^{-9}**
CD38	**4.16**	**2.21×10^{-8}**
***CD45* (*PTPRC*)**	**-1.64**	**1.89×10^{-8}**
***CD54* (*ICAM1*)**	**6.20**	**4.73×10^{-6}**
***CD20* (*MS4A1*)**	**-1.35**	**4.25×10^{-3}**
CD138 (*SDC1*)	-1.41	4.94×10^{-2}
CD126 (*IL6R*)	-1.12	8.67×10^{-2}
Be1 polarization marker		
IFNG	**-1.55**	**3.77×10^{-3}**
IL12RB1	-1.14	3.82×10^{-2}
IL12RB2	1.05	1.14×10^{-1}
IFNGR1	1.08	1.72×10^{-1}
IL12B	-1.05	2.04×10^{-1}
IL12A	1.01	3.17×10^{-1}
Be2 polarization marker		
IL4R	**-3.75**	**1.38×10^{-8}**
IL5RA	**-1.45**	**8.89×10^{-3}**
IL13	1.01	3.00×10^{-1}

Rows in bold type denote differentially expressed transcripts significant at an FDR < 1%. Be – B effector cell.

4.3.3 Differential expression of apoptotic markers by IFN-α

In light of the role of IFNs in the induction of programmed cell death during the course of viral infection, the effect of IFN-α stimulation on cell survival was assessed. The lymphocytic apoptotic markers *CD45* and *CD28* (Vigano et al., 2001) were down-regulated following IFN-α stimulation, at a significant and a suggestive level of significance, respectively (Table 4-5).

Furthermore, the transcripts for the anti-apoptotic markers c-Myc (*MYC*) and Bcl-2 (*BCL2*) were up-regulated on IFN-α stimulation (Hoffman and Liebermann, 2008). Transcripts were considered significantly differentially expressed at an FDR < 1%.

Table 4-5 The effect of IFN-α stimulation on the expression of pro-apoptotic and anti-apoptotic markers.

Apoptosis marker	Fold change	P_{FDR}
Pro-apoptotic markers		
***CD45* (*PTPRC*)**	**-1.64**	**1.89×10^{-8}**
CD28	-1.20	5.55×10^{-2}
Anti-apoptotic markers		
BCL2	**3.12**	**4.06×10^{-10}**
MYC	**1.64**	**1.61×10^{-3}**

Rows in bold type denote differentially expressed transcripts significant at an FDR < 1%.

4.3.4 Discussion

Acute IFN-α stimulation of B cells had a significant effect on global gene expression, including the induction of a type 1 IFN signature, and the differential expression of several B cell markers. The magnitude of effect of IFN-α stimulation reflects the importance of this cytokine in transcriptome modulation.

Stimulated B cells showed transcriptional changes indicative of antiviral response, immunoregulation, partial B cell activation and differentiation, and increased cell survival. This corroborates data from previous reports describing the downstream effects of IFN-α in murine and human B cells (Braun et al., 2002, Gujer et al., 2011, Apelbaum et al., 2013). Up-regulation of transcripts for the co-stimulatory molecules CD40 and CD80 is consistent with data from a mouse study (Braun et al., 2002), and supports at least partial activation of B cells by IFN-α stimulation. Failure to detect up-regulation of additional activation markers reported in the literature, such as *CD86*, *CD69*, and *CD27*, may be due to time point-dependent effects of the IFN-α response, heterogeneity of CD19+ B cell sub-populations, and differences between

murine and human biology. The cytokine-secreting role of B cells is effected through polarization into Be1 or Be2 cells, which secrete a panel of cytokines similar to that of T helper type 1 (Th1) and Th2 cells, respectively. IFN-α has been shown to trigger Be1 commitment after 18-24 hour stimulation of human naïve B cells, through increased production of IFN-γ and IL-12 (de Goer de Herve et al., 2011). The lack of change in gene expression patterns associated with Be1 (or Be2) polarization in the data presented here suggests that IFN-α priming for Be1 differentiation may occur after 6 hours of stimulation or, alternatively, may be concealed by the presence of multiple B cell sub-populations.

The effect of type 1 IFNs on cell survival is still unclear, with some studies reporting a role in induction of apoptosis, and others reporting proliferative effects and/or increased survival (Braun et al., 2002, Apelbaum et al., 2013). Type 1 IFN stimulation is generally insufficient to induce apoptosis in most cell types (Trinchieri, 2010), and is therefore likely to be dependent on cellular context. Tanaka and colleagues showed that type 1 IFN stimulation of murine primary embryonic fibroblasts selectively induced apoptosis in virally-infected cells, whilst promoting inhibition of viral replication and cell survival in uninfected cells (Tanaka et al., 1998). The data presented here show that B cells stimulated with IFN-α up-regulate pro-survival markers, but not apoptotic markers, further supporting the view that IFN-induced apoptosis in humans only occurs within a context of infection or prolonged IFN-α stimulation. Using evidence from previous studies reporting the effects of IFN-α in B cells, this work has shown that genome-wide microarray expression analysis may be useful in inferring the effect of cytokines in cell phenotype alterations.

4.4 Enrichment of disease-associated loci in the gene set differentially expressed by IFN-α

As summarised in Chapter 1, there is overwhelming evidence supporting a pathogenic role for IFN-α in a number of autoimmune diseases, particularly SLE, Sjögren's syndrome, and idiopathic inflammatory myopathies. Accordingly, an analysis was carried out to determine whether genes differentially expressed by IFN-α in B cells were enriched for complex disease-associated loci. Susceptibility genes for 13 complex diseases previously identified by genome-

wide association studies (GWAS) were tested for enrichment in the gene set differentially expressed by IFN-α. The diseases included eight autoimmune and five non-autoimmune diseases, of which a number have shown evidence of altered IFN activity. Disease-associated genes for SLE, rheumatoid arthritis, Sjögren's syndrome, idiopathic inflammatory myopathies, multiple sclerosis, inflammatory bowel disease (IBD), type 1 diabetes, psoriasis, osteoarthritis, type 2 diabetes, bipolar disorder, schizophrenia, and hypertension were curated from the National Human Genome Research Institute (NHGRI) catalogue of published GWAS (http://www.genome.gov/gwastudies/; dates accessed September 2014). The analysis included all disease-associated loci which passed a genome-wide significance threshold of 5×10^{-8}, and excluded intergenic and MHC loci (Appendix C). Section 4.5.5 describes the analysis of the extended MHC (xMHC) region.

Two datasets for SLE susceptibility genes were created: one curated from the NHGRI database for purposes of comparison with the other diseases (Appendix C, Table C1), and one derived from an updated PubMed search of all published SLE-associated genes (January 2016, Appendix D, Table D1). Over 40% of genes associated with SLE, rheumatoid arthritis, IBD, and multiple sclerosis were enriched in the gene set differentially expressed by IFN-α, at a significant (FDR < 5%) or suggestive level of significance (FDR < 7%, Table 4-6). In contrast, there was no significant enrichment for genes associated with psoriasis, type 1 diabetes, or the non-autoimmune diseases osteoarthritis, type 2 diabetes, bipolar disorder, schizophrenia, and hypertension. The small number of genes for Sjögren's syndrome and idiopathic inflammatory myopathies resulted in low power to yield any conclusive results.

Table 4-6 Enrichment analysis for loci associated with 13 complex diseases in the gene set differentially expressed by IFN-α (FDR < 1%).

Disease	Number of associated genes	Number of genes expressed in B cells	Percentage DE transcripts (DE/total)		$P_{enrichment}$	P_{FDR}
Autoimmune diseases						
Rheumatoid arthritis	**45**	**42**	**57.1%**	**(24/42)**	**3.83×10^{-5}**	**4.98×10^{-4}**
SLE (all published loci)	**83**	**74**	**43.0%**	**(34/79)**	**8.88×10^{-4}**	**5.77×10^{-3}**
IBD	*227*	*188*	*41.4%*	*(77/188)*	*2.57×10^{-2}*	*8.35×10^{-2}*
Multiple sclerosis	*63*	*56*	*40.0%*	*(22/56)*	*2.80×10^{-2}*	*8.35×10^{-2}*
SLE	*40*	*35*	*41.7%*	*(15/36)*	*3.21×10^{-2}*	*8.35×10^{-2}*
Psoriasis	20	15	46.7%	(7/15)	9.98×10^{-2}	0.22
Type 1 diabetes	60	43	30.2%	(13/43)	3.31×10^{-1}	0.54
Sjögren's syndrome	4	4	25.0%	(1/4)	1.00	1.00
IIM	1	1	100.0%	(1/1)	1.00	1.00
Non-autoimmune diseases						
Osteoarthritis	8	6	16.6%	(1/6)	4.96×10^{-1}	0.72
Type 2 diabetes	76	62	30.5%	(18/59)	2.73×10^{-1}	0.51
Bipolar disorder	9	5	20.0%	(1/5)	1.00	1.00
Schizophrenia	122	97	28.8%	(28/97)	7.52×10^{-1}	0.98
Hypertension	15	13	0.0%	(0/13)	-	-

Disease-associated loci were curated from the NHGRI GWAS database. An additional comprehensive gene list for SLE was also analysed (SLE (all published loci)). Disease enrichment was considered significant if $P < 0.05$ at an FDR < 5%. Significant p-values are displayed in bold type. P-values with suggestive significance ($P < 0.05$, FDR < 7%) are displayed in italic type. DE – differentially expressed, IBD – inflammatory bowel disease, SLE – systemic lupus erythematosus, IIM – idiopathic inflammatory myopathies.

4.4.1 Discussion

Disease enrichment analysis of genes differentially expressed by IFN-α in *ex vivo* B cells showed an overlap with susceptibility loci for the autoimmune diseases SLE, rheumatoid arthritis, IBD, and multiple sclerosis. This suggests a role for type 1 IFN in these diseases, which is discussed further below. Almost half of all psoriasis susceptibility genes also overlapped with the IFN-α data set at a non-significant p-value. In contrast, non-autoimmune

diseases, including type 2 diabetes, schizophrenia, bipolar disease, and hypertension, were not enriched. The small number of susceptibility genes for Sjögren's syndrome and idiopathic inflammatory myopathies available in the NHGRI data base resulted in a lack of power to detect disease enrichment in the IFN-α data set, given that these diseases also demonstrate a type 1 IFN signature (Mavragani et al., 2007, Lundberg and Helmers, 2010).

Administration of therapeutic type 1 IFN for hepatitis C and certain cancers has been reported to induce SLE, inflammatory arthritis, multiple sclerosis, psoriasis, and type 1 diabetes in a minority of patients (Crow, 2010, Afshar et al., 2013). A subset of patients with rheumatoid arthritis, and patients with relapsing remitting multiple sclerosis (RRMS), demonstrate a type 1 IFN signature, albeit less intense than that found in SLE (Crow, 2010). In contrast to SLE, both IFN-α and IFN-β have been found in the rheumatoid arthritis signature (Mavragani et al., 2010), and in brain lesions of RRMS patients (Traugott and Lebon, 1988). IFN-β was proposed to have a regulatory role in multiple sclerosis, which led to the administration of this cytokine as a therapeutic agent in patients (Buttmann and Rieckmann, 2007). Similarly, IFN-β is associated with disease improvement in the joints of murine arthritis models and rheumatoid arthritis patients, further supporting a protective role for this cytokine (Crow, 2010). The suggestive enrichment of rheumatoid arthritis and multiple sclerosis candidate genes following IFN-α stimulation may indicate a role for IFN-α in these diseases, but may also reflect the redundancy of the IFN-α and IFN-β signalling pathways.

Interestingly, type 1 diabetes susceptibility genes were not enriched in the IFN-α data set. The type 1 IFN signature is not a hallmark of type 1 diabetes, although a recent study reported presence of the signature in peripheral blood preceding the onset of type 1 diabetes (Ferreira et al., 2014). Type 1 diabetes patients given low-dose IFN-α (5000 units) demonstrate less severe disease than patients given a placebo or high-dose IFN-α (30,000 units, Rother et al., 2009). Altogether, these observations support a small role for this cytokine in the pathogenesis of type 1 diabetes in B cells. Studies in other relevant cell types for type 1 diabetes, such as pancreatic β cells, may help to clarify the role of IFN-α in this disease. Similarly, the effect of type 1 IFNs in IBD is controversial, with one report claiming both regulatory and inflammatory roles in different phases of the disease (Rauch et al., 2014). Finally, therapeutic IFN-α administration has also been reported to trigger or exacerbate neuropsychiatric symptoms, including bipolar disorder

and depression (Mustafa et al., 2014). Various studies have shown a decreased IFN-α response in schizophrenia patients; however, the role of IFN-α in schizophrenia is still unclear (Inglot et al., 1994, Hornberg et al., 1995, Katila et al., 1989). The data presented in this chapter suggest that genetic susceptibility loci for both schizophrenia and bipolar disease may not interact directly with the IFN-α pathway, such that alterations of IFN-α levels in these patients may result from an epiphenomenon.

As loci identified by GWAS are often located in non-coding genic or intergenic regions of the genome, gene annotations from the NHGRI database have been established according to the gene most proximal to the associated marker. The annotated genes in the NHGRI data base may not correspond to actual disease-causing genes, as the molecular mechanisms underlying disease associations may involve loci located in proximity (*cis*) or distant (*trans*) to the associated marker. A causal regulatory locus can act on a different locus through expression quantitative trait loci (eQTLs), epigenetic changes, chromatin conformation changes, or by affecting microRNA sites (Pai et al., 2015, Voisin et al., 2015). Some of the disease-associated loci described in this chapter may therefore not be biologically relevant in SLE. Since GWAS do not reveal the functional context for genetic associations, the information gained from GWAS needs to be combined with regulatory and functional data generated in disease-relevant conditions, such as the gene expression data here described. This will enable fine-mapping and identification of the causal genes and functional mechanisms underlying disease associations.

In summary, the work described here shows that IFN-α stimulation of B cells regulates the expression of candidate genes for SLE and rheumatoid arthritis, and possibly candidate genes for multiple sclerosis and IBD, implicating a role for this cytokine in these autoimmune diseases. Since few studies have addressed the role of type 1 IFN in perpetuating inflammation and tissue destruction in autoimmune diseases other than SLE, these data provide a platform for further investigation.

4.5 Differential expression of genes associated with idiopathic and monogenic lupus by IFN-α

4.5.1 Idiopathic SLE susceptibility genes

Based on the comprehensive PubMed literature search described in section 4.4, as of January 2016, 83 genes were associated with susceptibility to idiopathic SLE (using a statistical threshold of 5×10^{-8}, Appendix D, Table D1). Of these 83 genes, 74 transcripts were expressed in the *ex vivo* B cell data set studied. However, the mechanisms underlying their association with disease remain largely unknown. I therefore sought to explore how the expression of SLE susceptibility genes is regulated by IFN-α stimulation in B cells (Table 4-7). As expected, significantly differentially expressed SLE candidate genes included those functionally involved in canonical Toll-like receptor (TLR) and type 1 IFN pathways, such as *FCGR2B* and *IFIH1* (MDA5) receptors, the signalling molecules *STAT4* and *TYK2*, and the IFN inducible genes *IRF5* and *IRF7*. Several of the genes differentially expressed by IFN-α encode BCR signalling components, including *BANK1*, *BLK*, *LYN*, and *PRKCB*, which are known to cross-talk with the type 1 IFN pathway (Delgado-Vega et al., 2010).

Curiously, a number of SLE candidate genes outside canonical type 1 IFN and BCR signalling pathways (henceforth referred to as non-canonical genes) were also differentially expressed following IFN-α stimulation. These include moderately expressed genes (mean $\log_2$ expression values between 4 and 5) whose function in B cells is unknown or poorly characterised (e.g. *NCF2*, *FCGR2A*, and *ITGAM*), and genes of unknown function (e.g. *PXK* and *TMEM39A*). Differential expression of two canonical and six non-canonical SLE candidate genes was validated by qPCR, using cDNA from the exon array samples (FDR < 1%, Table 4-8). These eight genes were selected for qPCR validation, given the ongoing functional studies on their role in SLE pathogenesis.

Table 4-7 The effect of IFN-α stimulation on the expression of genes associated with idiopathic SLE.

Differentially expressed transcript	Transcript cluster ID	Fold change	P_{FDR}	Role
Strong evidence of DE (FDR < 1%)				
IFIH1 (MDA5)	2584207	9.82	6.09×10^{-16}	Type 1 IFN pathway
FLI1	3355733	-3.33	1.03×10^{-14}	Lymphoid Txn Factor
CIITA	3647993	-3.23	1.12×10^{-13}	Transcriptional activator of HLA class I and class II
IKZF1	3001479	-4.25	2.88×10^{-13}	Lymphocyte development
IRF7	3358174	5.31	5.10×10^{-13}	Type 1 IFN pathway
SMG7	2371255	2.30	5.15×10^{-13}	Nonsense-mediated mRNA decay
TLR7	3969081	8.41	4.91×10^{-12}	Type 1 IFN pathway
STAT4	2592356	7.24	1.31×10^{-11}	Type 1 IFN pathway
TNPO3	3072014	1.98	1.64×10^{-10}	Nuclear import receptor
BLK	3085990	-2.69	3.07×10^{-10}	BCR signalling
SH2B3	3431892	2.78	6.27×10^{-10}	Hematopoiesis
NADSYN1	3338968	-1.66	1.17×10^{-9}	Metabolic redox reactions
PRKCB	3653123	-3.61	2.23×10^{-9}	BCR signalling
ELF1	3511031	1.73	3.66×10^{-9}	Lymphoid Txn Factor
SOCS1	3680213	3.96	4.92×10^{-9}	Cytokine signalling regulator
CSK	3601840	-1.73	1.26×10^{-8}	BCR signalling
CXorf21	4003895	2.41	3.69×10^{-8}	Unknown
ITGAM	3656990	-2.80	5.24×10^{-8}	Phagocytosis in neutrophils and monocytes/macrophages
PTPN22	2428796	-2.15	8.89×10^{-8}	Lymphoid signalling
IKZF2	2597867	2.32	9.06×10^{-8}	Lymphocyte development
TREX1	2621705	1.48	1.23×10^{-7}	Nucleic acid metabolism
PLD2	3707214	-1.44	3.46×10^{-7}	Phospholipase
FCGR2B	2363808	-2.54	4.12×10^{-7}	B cell Fc receptor (inhibitory)
BANK1	2737596	-1.83	6.76×10^{-7}	BCR signalling
RASGRP3	2476671	1.72	1.83×10^{-6}	Energy metabolism
KIAA0319L	2406139	-1.44	2.14×10^{-6}	Unknown
FCGR2A	2363689	-3.38	2.87×10^{-6}	Monocyte/macrophage Fc receptor (activating)

Differentially expressed transcript	Transcript cluster ID	Fold change	P_{FDR}	Role
ICAM1	3820443	6.20	4.73×10^{-6}	Cell adhesion
ETS1	3397589	-1.80	4.91×10^{-6}	Type 1 IFN response inhibitor
NCF2	2447414	-1.78	9.20×10^{-6}	Phagosome function in neutrophils
PXK	2626167	-1.69	2.92×10^{-5}	Unknown
TMEM39A	2690850	1.36	2.95×10^{-5}	Unknown
LYN	3098977	1.49	1.27×10^{-4}	BCR signalling
CDKN1B	3405440	-1.31	1.45×10^{-4}	Cell cycle
MECP2	4027056	1.62	1.96×10^{-4}	DNA methylation
DHCR7	3380697	-1.45	4.21×10^{-4}	Cholesterol production
FCGR3B	2440943	-1.94	1.20×10^{-3}	NK cell receptor
RAD51B	3541497	-1.25	1.90×10^{-3}	DNA repair
TCF7	2829171	-1.41	2.79×10^{-3}	Lymphocyte differentiation
TNFSF4	2444283	1.63	3.71×10^{-3}	BCR signalling
TYK2	3850278	-1.16	3.87×10^{-3}	Type 1 IFN pathway
CD80	2690900	1.47	5.37×10^{-3}	BCR signalling
TNFAIP3	2927506	1.93	6.25×10^{-3}	Type 1 IFN pathway
IRF5	3023246	1.29	7.05×10^{-3}	Type 1 IFN pathway
UHRF1BP1	2904270	-1.11	7.39×10^{-3}	Unknown
UBE2L3	3938175	-1.25	9.78×10^{-3}	NFκB pathway
Suggestive evidence of DE (FDR < 5%)				
SPRED2	2556752	-1.18	1.02×10^{-2}	Growth factor signalling
SKP1	2876011	-1.32	1.15×10^{-2}	Ubiquitination
TET3	2489071	1.24	1.19×10^{-2}	DNA methylation
ZPBP2	3720543	-1.21	1.38×10^{-2}	Unknown
ATG5	2967550	1.15	1.77×10^{-2}	Autophagy
DRAM1	3428783	1.28	2.30×10^{-2}	NFκB pathway and autophagy
IRF8	3672489	1.25	3.81×10^{-2}	Type 1 IFN pathway
IL21	2784279	-1.21	4.12×10^{-2}	Immune cell differentiation and activity

DE – differential expression, BCR – B cell receptor, Txn Factor – transcription factor, NK – natural killer, NFκB – nuclear factor κB.

Table 4-8 qPCR validation of differentially expressed idiopathic SLE susceptibility genes on IFN-α stimulation.

Differentially expressed transcript	Fold change	P_{FDR}
IFIH1 (MDA5)	10.96	6.58×10^{-20}
ITGAM	-7.95	4.03×10^{-18}
IRF7	40.53	2.23×10^{-17}
IKZF1	-3.35	1.18×10^{-16}
FCGR2A	-9.65	1.02×10^{-9}
PTPN22	-2.25	1.02×10^{-9}
TNPO3	1.50	1.85×10^{-8}
FCGR3B	-2.97	5.41×10^{-5}

Transcripts were considered significantly differentially expressed at a FDR < 1%.

4.5.2 Monogenic lupus susceptibility genes

An investigation of the effect of IFN-α on genes that cause rare monogenic forms of lupus was carried out. Eighteen genes associated with monogenic lupus were identified from a PubMed literature search (Appendix D, Table D2), of which 16 transcripts were expressed in the B cell expression data set. Seven monogenic lupus genes (46%) showed strong or suggestive evidence of differential expression by IFN-α (FDR < 1% or < 5%, respectively, Table 4-9). These included genes involved in nucleic acid metabolism (*SAMHD1*, *TREX1*, *DNASE1L3*), the complement components *C1R*, *C1S* and *C2*, and acid phosphatase 5 (*ACP5* or *TRAP*). In addition, differential expression of *PRKCD* (P_{FDR} = 0.08) and *RNASEH2C* (P_{FDR} = 0.06) demonstrated a trend towards significance. Differential expression five transcripts, including *PRKCD* and *RNASEH2C*, was validated by qPCR, using cDNA from the exon array samples (Table 4-10).

Table 4-9 The effect of IFN-α stimulation on the expression of genes associated with monogenic lupus.

Differentially expressed transcript	Transcript cluster ID	Exon array		Role
		Fold change	P_{FDR}	
Strong evidence of DE (FDR < 1%)				
SAMHD1	3904691	3.76	2.14×10^{-8}	Nucleic acid metabolism
TREX1	2621705	1.48	1.23×10^{-7}	Nucleic acid metabolism
DNASE1L3	2678298	-1.93	1.09×10^{-4}	Nucleic acid metabolism
C1R	3442475	-1.22	3.58×10^{-4}	Complement
Suggestive evidence of DE (FDR < 5%)				
C1S	3403168	-1.21	1.12×10^{-2}	Complement
ACP5 (TRAP)	3851072	2.01	2.56×10^{-2}	Type 1 IFN regulator
C2	2902804	-1.22	3.00×10^{-2}	Complement
No evidence of DE (FDR > 5%)				
PRKCD	2624291	-1.12	8.03×10^{-2}	Apoptosis of B cells
RNASEH2C	3377826	1.14	6.80×10^{-2}	Nucleic acid metabolism

DE – differential expression.

Table 4-10 qPCR validation of differentially expressed monogenic lupus susceptibility genes on IFN-α stimulation.

Differentially expressed transcript	qPCR	
	Fold change	P_{FDR}
Strong evidence of DE (FDR < 1%)		
SAMHD1	6.01	5.26×10^{-15}
PRKCD	-1.31	2.20×10^{-4}
ACP5 (TRAP)	1.61	4.84×10^{-4}
DNASE1L3	-2.93	1.12×10^{-3}
Suggestive evidence of DE (FDR < 5%)		
RNASEH2C	1.10	1.87×10^{-2}

DE – differential expression.

4.5.3 IFN-α effects mirror the functional consequences of SLE polymorphisms

Several studies have investigated the functional consequences of established SLE-associated polymorphisms, such as differential transcript and protein expression, or altered cellular function. Interestingly, the direction of effect of IFN-α on transcript levels (increasing or decreasing) of several SLE susceptibility genes mirrored the functional effects of gain-of-function (GoF) and loss-of-function (LoF) SLE polymorphisms reported in the literature (summarized in Table 4-11). Differentially expressed SLE genes within canonical type 1 IFN and BCR pathways which harbour GoF (*IFIH1*, *IRF7*, *STAT4*, *TNFSF4*) and LoF (*BLK*, *FCGR2B*, *TYK2*) polymorphisms were respectively up- and down-regulated by IFN-α. The same pattern was observed for nine non-canonical SLE genes, and five monogenic lupus-associated genes (Table 4-12).

Table 4-11 Differential expression of canonical and non-canonical SLE-associated genes by IFN-α mirrors the effect of GoF and LoF polymorphisms within those genes.

Transcript	IFN-α-induced effect on gene expression	SLE-associated polymorphism	Functional effect of polymorphism	Reference
Canonical TLR, type 1 IFN, and BCR signalling SLE genes				
IFIH1 (MDA5)	⬆	rs1990760 (A946T)	Increased sensitivity to IFN-α in SLE PBMCs	Robinson et al. (2011)
IRF7	⬆	rs1131665 (Q412R)	Elevated IRF7 activity in reporter cells	Fu et al. (2011)
STAT4	⬆	rs7582694	Increased *STAT4* expression in healthy PBMCs	Abelson et al. (2009)
BLK	⬇	rs13277113, rs2736340	Reduced *BLK* expression in healthy B cells	Hom et al. (2008), Bentham et al. (2015)
FCGR2B	⬇	rs1801274, I232T	Reduced *FCGR2B* expression in healthy B cells; protein LoF	Bentham et al. (2015), Floto et al. (2005)
TNFSF4	⬆	rs2205960	Increased *TNFSF4* expression in SLE blood lymphocytes	Cunninghame Graham et al. (2008), Manku et al. (2013)
TYK2	⬇	rs2304256	Reduced *TYK2* expression in healthy monocytes	Bentham et al. (2015)
IRF5	⬆	rs10488631	Increased *IRF5* transcript and protein expression in SLE PBMCs	Feng et al. (2010)
Non-canonical SLE genes				
SOCS1	⬆	rs9652601	Increased *SOCS1* expression in healthy B cells	Bentham et al. (2015)
ITGAM	⬇	R77H	Protein LoF	Rhodes et al. (2012)
PTPN22	⬇	R620W	Conflicting	Holmes et al. (2015), Wang et al. (2015)
FCGR2A	⬇	H166R, R131H	Reduced protein function	Tada et al. (2012), Brown et al. (2007)
ETS1	⬇	rs1128334, rs6590330	Reduced *ETS1* expression	Lu et al. (2015), Yang et al. (2010)
NCF2	⬇	H389Q, R395W	Protein LoF	Jacob et al. (2012)
DHCR7	⬇	rs3794060	Reduced *DHCR7* expression in healthy stimulated monocytes	Bentham et al. (2015)
FCGR3B	⬇	CNV	Reduced protein expression	Willcocks et al. (2008)
SKP1	⬇	rs7726414	Reduced *SKP1* expression in healthy B cells	Bentham et al. (2015)

GoF – gain of function, LoF – loss of function, PBMCs – peripheral blood mononuclear cells, CNV – copy number variant.

Table 4-12 Differential expression of monogenic lupus-associated genes by IFN-α mirrors the effect of LoF polymorphisms within those genes.

Transcript	IFN-α-induced effect on gene expression	SLE-associated polymorphism	Functional effect of polymorphism	Reference
Monogenic lupus genes				
DNASE1L3	⬇	W215GfsX2	Protein LoF	Al-Mayouf et al. (2011)
C1R	⬇	Homozygous deficiency	Protein deficiency	Bryan and Wu (2014)
C1S	⬇	Homozygous deficiency	Protein deficiency	Bryan and Wu (2014)
C2	⬇	Homozygous deficiency	Protein deficiency	Bryan and Wu (2014)
PRKCD	⬇	G510S	Reduced *PRKCD* expression and activity	Belot et al. (2013)

LoF – loss of function.

4.5.4 Type 1 interferonopathy susceptibility genes

Genetic studies of type 1 interferonopathies, which comprise clinical syndromes characterized by high levels of type 1 IFN, have advanced our understanding of the molecular pathways involved in triggering type 1 IFN production. Studies in patients with Aicardi-Goutières syndrome, a neurodegenerative type 1 interferonopathy, have identified LoF mutations in genes involved in nucleic acid metabolism and recognition (such as *TREX1*, *SAMHD1*, *RNASEH2C*, and *IFIH1*), which correlate with increased serum type 1 IFN levels (Miner and Diamond, 2014, Kretschmer et al., 2015, Gunther et al., 2015). Gene expression data showed that IFN-α stimulation resulted in the up-regulation of eight out of 13 genes associated with the type 1 interferonopathies (FDR < 5%, Appendix E). Interestingly, five of the up-regulated genes overlapped with idiopathic and monogenic lupus susceptibility genes (Figure 4-5). Out of a total of 33 differentially expressed loci from the three disease groups, *TREX1* is the only gene shared by all.

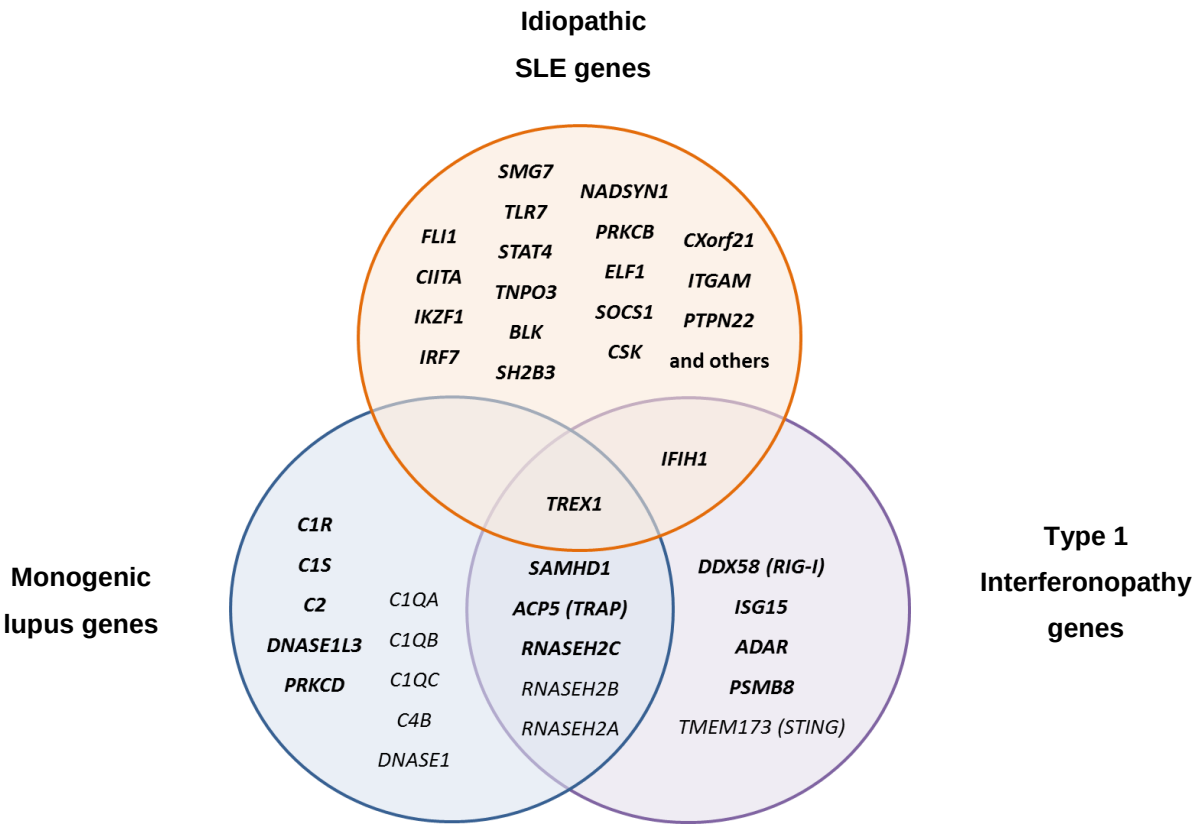


Figure 4-5 Venn diagram showing differentially expressed susceptibility genes for lupus and type 1 interferonopathies following IFN-α stimulation.

The diagram demonstrates shared and unique susceptibility loci for idiopathic SLE (top 21 differentially expressed genes shown), monogenic lupus, and type 1 interferonopathies. Genes differentially expressed by IFN-α at an FDR < 5% are shown in bold type.

4.5.5 xMHC region genes

The strongest genetic risk for SLE in Europeans arises from the MHC region, with a substantial contribution from extended ancestral human leukocyte antigen (HLA) haplotypes. Given the extent of linkage disequilibrium (LD) in the region, and the challenges this poses for fine-mapping disease associations, the effect of IFN-α on gene expression within the xMHC region was examined separately from that on SLE-associated genes outside the xMHC.

IFN-α significantly affected the expression of 56% of xMHC genes (110 out of 197 genes, FDR < 1%, Figure 4-3), which are listed in Appendix F, Table F1. The most significantly up-regulated genes were located within the class III region, and include tumour necrosis factor (TNF) family members (*TNF*, *LTA*), antigen processing genes for HLA class I presentation (*TAP1*, *TAP2*, *PSMB8*), and the heat shock protein gene *HSPA1B*. Owing to the highly polymorphic nature of

HLA genes, many of the exon array probes targeting these genes were removed from the data during QC. Excluded probes included those for the classical class I *HLA-A*, *HLA-B*, *HLA-C*, and classical class II *HLA-DRB* and *HLA-DPB1* transcripts. Nevertheless, IFN-α up-regulated two non-classical HLA class I transcripts (*HLA-E* and *HLA-F*). In contrast, HLA class II transcripts were down-regulated, including classical antigen presenting molecules (*HLA-DRA*, *HLA-DPA1*), non-classical HLA molecules involved in peptide loading (*HLA-DMA*, *HLA-DMB*, *HLA-DOA*, *HLA-DOB*), and the pseudogene *HLA-DPB2*. Down-regulation of the classical class II gene *DPB1* was demonstrated by qPCR (P_{FDR} = 4.05 × 10^{-3}, FC = -1.67). Additional genes down-regulated by IFN-α in the exon array data included several members of the histone family located in the xMHC class I (14 out of 36 histone genes).

Several of the xMHC genes differentially expressed by IFN-α stimulation have been associated with SLE in GWAS and candidate locus studies (as detailed in Chapter 3, section 3.1, and Chapter 5, section 5.1). These included genes which have shown single nucleotide polymorphism (SNP) associations with SLE independently of extended ancestral HLA haplotypes (*MICB*, *MSH5*, and *DPB1*), and genes in strong LD with extended ancestral HLA haplotypes, such as *HLA-DRB1*03:01* (*NCR3*, *AIF1*, *TNF*, *LTA*, *LTB*, *C6orf10*, *TRIM27*, and *HSD17B8*, FDR < 1%).

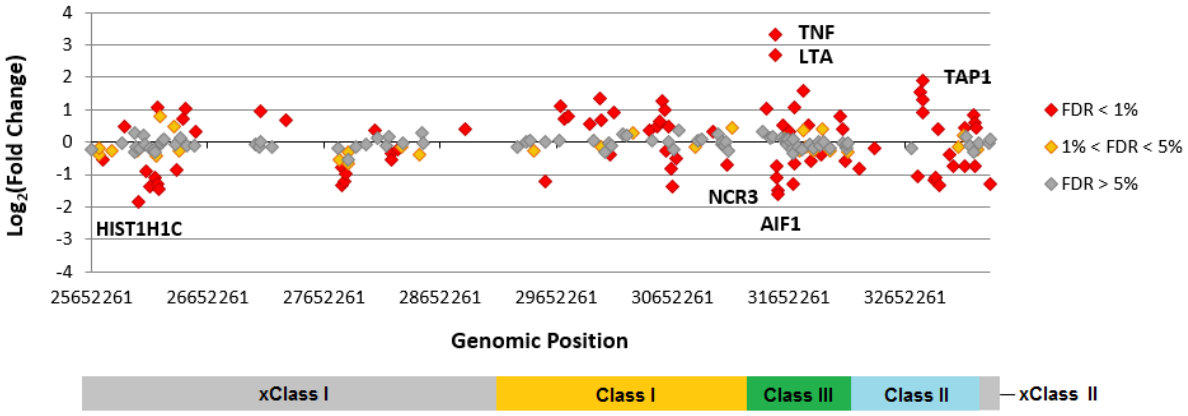


Figure 4-6 Effect of IFN-α stimulation on differential gene expression in the xMHC region.

The plot shows that genes with the largest fold changes are located within the MHC class III region. Genes with the smallest fold changes are located within the xMHC class I and MHC class III regions. The top three up- and down-regulated genes are labelled.

4.5.6 Discussion

Investigation of SLE candidate genes differentially expressed by IFN-α revealed that multiple transcripts were positively or negatively regulated in *ex vivo* B cells. Differentially expressed genes included both canonical and non-canonical SLE candidate genes, and genes within the xMHC region. This suggests a direct role for these genes in the pathogenesis of SLE through the activation of the type 1 IFN pathway. Furthermore, the effect of IFN-α on the expression of several idiopathic and monogenic lupus candidate genes mirrors the genetic consequences of risk polymorphisms reported in functional studies.

As expected, SLE susceptibility genes differentially expressed by IFN-α included those functionally involved in canonical TLR, type 1 IFN, and BCR signalling pathways. Up-regulation of TLR pathway genes indicate that IFN-α might possibly induce a positive feedback loop in type 1 IFN production. Differential expression of BCR signalling pathway genes was likely to result from the crosstalk between type 1 IFN signalling and BCR signalling, as previously mentioned. Functional studies on the IFN signalling genes *STAT4*, *IRF5*, *IRF7*, and *IFIH1* have demonstrated that SLE risk GoF polymorphisms correlate with increased transcript expression or protein activity, as well as an exacerbated response to IFN-α (Abelson et al., 2009, Kariuki et al., 2009, Feng et al., 2010, Niewold et al., 2008, Fu et al., 2011, Salloum et al., 2010, Robinson et al., 2011). Up-regulation of these genes by IFN-α mirrors the genetic effects of functional polymorphisms, and again suggests a positive feedback loop induced by IFN-α. Similarly, down-regulation of the canonical genes *BLK* and *FCGR2B* by IFN-α parallels the effect of reported LoF polymorphisms in these genes. Furthermore, SLE-associated LoF polymorphisms in *FCGR2B* have been shown to result in increased pro-inflammatory signalling (Floto et al., 2005). These canonical SLE candidate genes may therefore play a role not only in inducing the type 1 IFN response, but are also likely to perpetuate it via an IFN-α positive feedback loop.

Many of the differentially expressed idiopathic and monogenic SLE candidate genes were not previously known to be functionally associated with the type 1 IFN or BCR signalling pathways following stimulation with IFN-α (i.e. non-canonical genes). Interestingly, these included SLE genes whose function in B cells is poorly characterised or unknown, such as *ITGAM*, *FCGR2A*, *FCGR3B*, and *NCF2*, as well as genes conventionally expressed in B cell, such as *ETS1*, *PTPN22*, and *IKZF1*. Again, down-regulation of several of these genes by IFN-α parallels the

functional effects of SLE risk polymorphisms. The implications of these observations for the pathogenesis of SLE are discussed further below, and are illustrated in Figure 4-7 and Figure 4-8 for a number of SLE candidate genes.

ITGAM encodes a complement component 3 receptor subunit, CD11b, which acts as an anti-inflammatory mediator by inducing phagocytosis of immune complexes (ICs). Recently, an SLE-associated LoF polymorphism in *ITGAM* was shown to decrease the phagocytic activity of macrophages, which is thought to affect the uptake of ICs (Rhodes et al., 2012). CD11b is abundantly expressed in monocytes and neutrophils; however, its expression and function in human B cells is not well characterized, except in B1 and memory (CD27+) B cell sub-populations (Griffin et al., 2012, Kawai et al., 2005). The work here described confirms moderate expression of *ITGAM* in CD19+ B cells, and reports, for the first time, down-regulation of *ITGAM* transcripts by IFN-α stimulation. Interestingly, CD11b expression is reduced in T cells, monocytes, and neutrophils from SLE patients (Zhou et al., 2014). Furthermore, CD11b was recently shown to negatively regulate the survival of murine autoreactive B cells (Ding et al., 2013). Based on these data, it is possible to hypothesise that IFN-α increases the survival of autoreactive B cells through the down-regulation of *ITGAM*, thereby predisposing to autoimmunity. However, one study reported that type 1 IFN stimulation mediated conversion of plasmacytoid DCs (pDCs) into CD11b+ DCs (Liou et al., 2008), implying an increase in *ITGAM* expression in pDCs. The effect of IFN-α on *ITGAM* expression may therefore be cell type-dependant.

Other differentially expressed SLE candidate genes poorly characterized in B cells include the Fcγ receptors *FCGR2A* and *FCGR3B*. These are activating receptors which induce endocytosis of ICs. *FCGR2A* is mainly expressed in phagocytic cells (macrophages, neutrophils, and pDCs), and harbours LoF polymorphisms associated with SLE (Tada et al., 2012). The copy number variable gene *FCGR3B* is constitutively expressed in neutrophils, and low copy number is associated with risk of SLE (Yuan et al., 2015). Functional studies of the risk polymorphisms in both *FCGR2A* and *FCGR3B* have shown that cells exhibit decreased IC clearance (Deng and Tsao, 2010). This chapter reports a decrease in expression of both *FCGR2A* and *FCGR3B* transcripts by IFN-α, which may translate to decreased activity of these receptors in B cells. However, *FCGR3B* mRNA expression in B cells has not been previously confirmed in the

literature, due to high sequence similarity within copy number variable loci, and flow cytometric data have suggested that it is not expressed at protein-level in B cells (Dr Ruth Tarzi, Imperial College London; personal communication). In order to validate the results presented here, protein-level expression of *FCGR3B* in B cells would need to be demonstrated. Conversely, transcriptional changes may not be translated at protein-level, due to additional regulatory mechanisms, such as nonsense-mediated mRNA decay or translational repression.

ETS1 encodes a transcription factor highly expressed in lymphoid tissues, which binds the IFN-stimulated response element (ISRE), and acts as a repressor of IFN-α production and B cell development (Flesher et al., 2010). SLE-associated polymorphisms in *ETS1* correlate with decreased mRNA expression and increased IFN signalling in Asian populations (Lu et al., 2015, Yang et al., 2010). Furthermore, *ETS1* transcript levels are decreased in peripheral blood mononuclear cells (PBMCs) from SLE patients compared to controls (Li et al., 2010). The data presented here showed that IFN-α stimulation also down-regulated *ETS1* expression, which suggests that reduced levels of this transcript in SLE patients may be due to environmental as well as genetic factors.

PTPN22 encodes a lymphoid tyrosine phosphatase, Lyp, which is involved in immune cell development and function, and is an inhibitor of immune cell signalling (Bottini and Peterson, 2014). The non-synonymous polymorphism R620W is associated with SLE, as well as various other autoimmune diseases (Gregersen et al., 2012). Functional studies have shown conflicting results regarding the effect of R620W, with reports showing both an increase (due to LoF) and a decrease (due to GoF) in T cell and B cell signalling. A study by Holmes and colleagues (Holmes et al., 2015) showed that Lyp-deficient mice exhibited increased serum IFN-α, as did SLE patients carrying the risk polymorphism. The study concluded that Lyp negatively regulates IFN-α signalling, and that the *PTPN22* risk variant results in LoF. In contrast, a different group demonstrated that SLE patients carrying the R620W variant show reduced IFN-α signalling in PBMCs (Wang et al., 2015). The effect of this *PTPN22* polymorphism may be dependent on cell sub-populations, specific immune pathways, or biological state (Burn et al., 2011, Bottini and Peterson, 2014). The data presented in this chapter showed that IFN-α down-regulated *PTPN22* in B cells, compounding the observations of Holmes and colleagues (Holmes et al.,

2015), and may therefore help to clarify the functional role of this archetypal autoimmune gene within the context of B cells.

IKZF1 encodes a member of the Ikaros family of transcription factors (Ikaros), which is involved in lymphocyte differentiation (Hu et al., 2013). Reports of the functional effect of the SLE-associated SNP, rs4917014, on *IKZF1* expression levels are contradictory. One study reported a correlation between the risk allele and up-regulation of *IKZF1* transcripts *in cis,* and up-regulation of IFN-α-inducible transcripts *in trans,* in peripheral blood cells (Westra et al., 2015). However, the *cis*-eQTL has not been replicated in other gene expression cohorts (Bentham et al., 2015). The down-regulation of *IKZF1* by IFN-α here reported is of interest, given that *IKZF1* is also down-regulated in the serum of SLE patients (Costa-Reis and Sullivan, 2013). It has been suggested that low levels of *IKZF1* expression activate IFN-α production through regulation of *STAT4* transcription (Dang et al., 2014). In contrast, the data presented here showed that the SLE-associated Ikaros family members *IKZF2* (Helios) and *IKZF3* (Aiolos) were not down-regulated by IFN-α. Taken together, these observations support a direct relationship between levels of *IKZF1* in SLE and IFN-α activity, and suggest that *IKZF1* may play a role in a negative feedback loop in the type 1 IFN response.

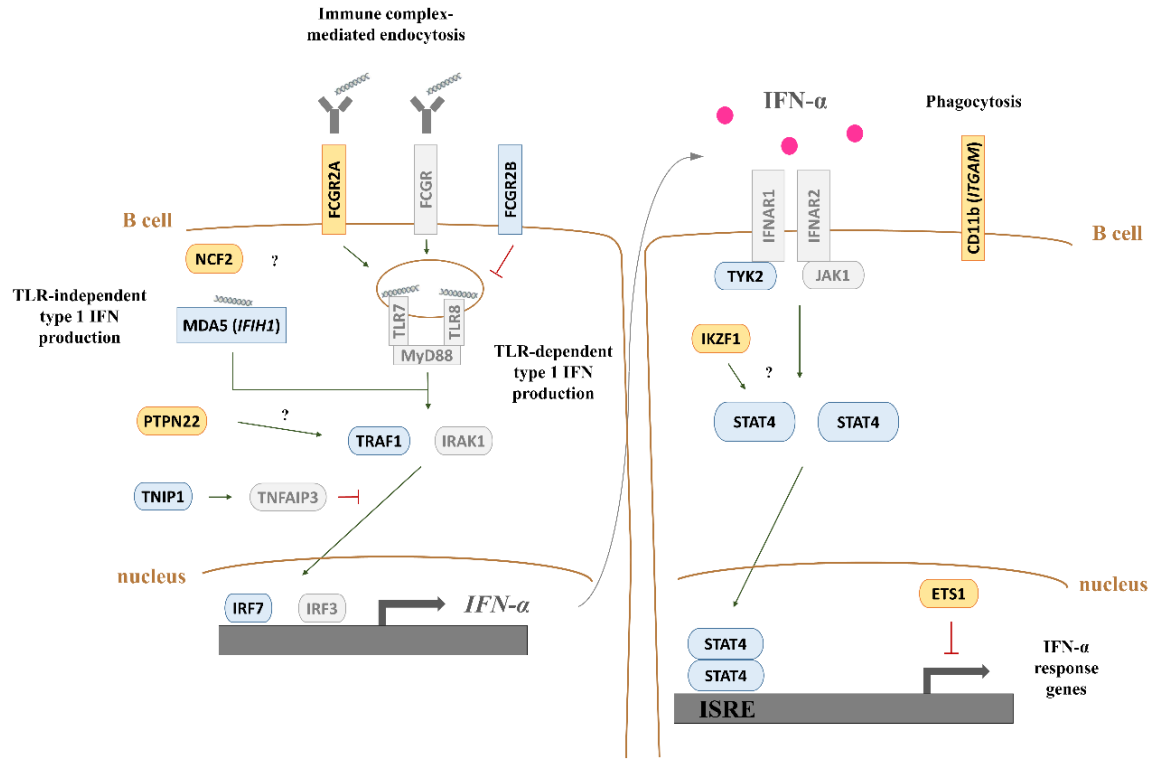


Figure 4-7 Differential expression of genes associated with idiopathic SLE by IFN-α stimulation.

SLE candidate genes within canonical TLR and type 1 IFN pathways are coloured in blue and those within non-canonical pathways are coloured in orange. Non-differentially expressed genes relevant to the pathways are coloured in grey. The activating and inhibitory effects of SLE genes on IFN-α production are illustrated by green and red arrows, respectively.

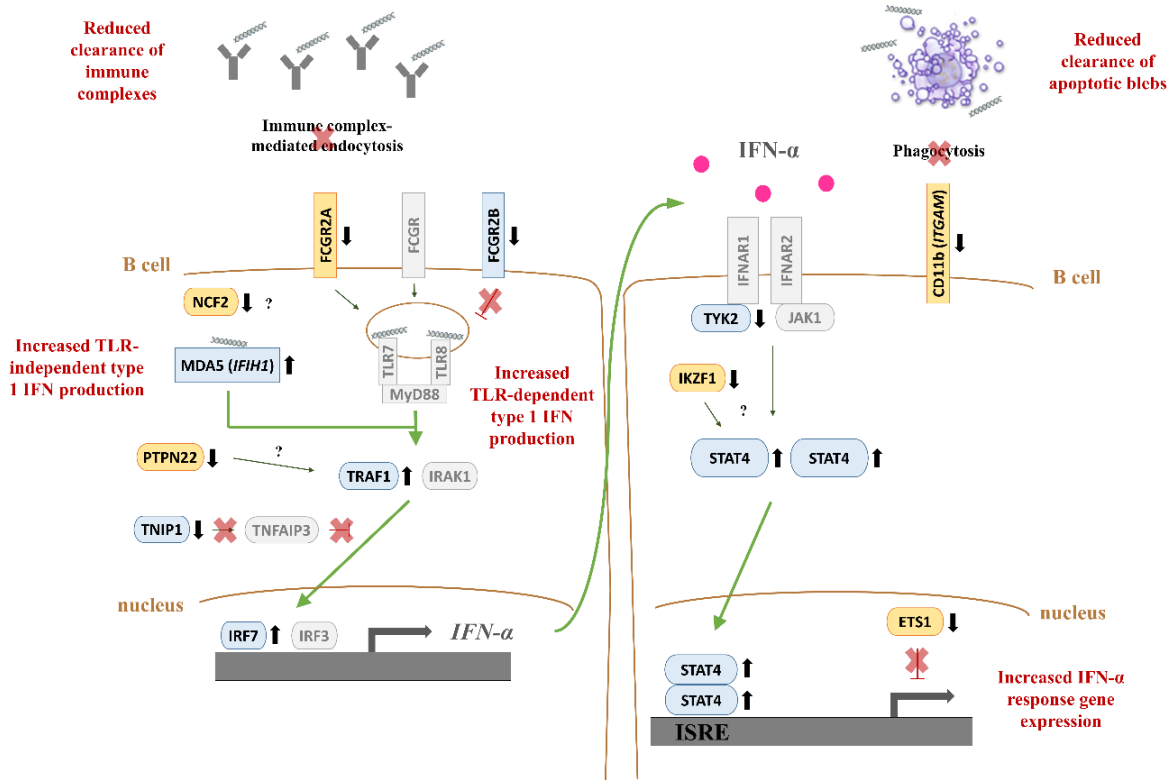


Figure 4-8 Potential consequences of IFN-α on the function of genes associated with idiopathic SLE.

Associated genes within canonical TLR and type 1 IFN pathways are coloured in blue and those within non-canonical pathways are coloured in orange. Non-differentially expressed genes relevant to the pathways are coloured in grey. Black arrows indicate the direction of effect on gene expression levels following IFN-α. The resulting hypothetical increase and decrease in protein function is indicated by thick green arrows and red crosses, respectively. Changes in the expression of these genes may increase the IFN-α response through reduced clearance of immune complexes and apoptotic debris, increased TLR-dependent and independent IFN-α production, and increased type 1 IFN signalling.

Monogenic (or familial) lupus comprise rare inherited forms of lupus caused by highly penetrant, single-gene mutations. IFN-α induced differential expression of a range of monogenic lupus-associated genes, including nucleic acid metabolism genes, complement components, and type 1 IFN signalling genes (Figure 4-9). IFN-α had a minimal negative effect on complement expression, which had been previously demonstrated in PBMCs (Colten et al., 1986, Sánchez-Pernaute et al., 2015). Deficiencies in components of the early classical complement pathway, including *C1Q*, *C1R*, *C1S*, and *C2*, have long been associated with an increased risk of lupus (Bryan and Wu, 2014). However, due to the high LD between *C2* and other MHC genes, independence of this signal from HLA haplotypes has not been confirmed. Complement deficiency results in impaired clearance of ICs and apoptotic cells, which provide a source of self-antigens to autoreactive cells, and activate type 1 IFN production (Belot and Cimaz, 2012). One can hypothesise that down-regulation of complement components by IFN-α, combined with a genetically-driven decrease in immune regulation, may result in the breakdown of tolerance and transition to autoimmunity (Figure 4-10).

Similarly, several LoF mutations in the nuclease genes *TREX1*, *DNASE1L3*, and *RNASEH2C*, and in the genome-stabilising gene *SAMHD1*, have been shown to impair clearance of intracellular nucleic acids and induce type 1 IFN activation (Fye et al., 2011, Al-Mayouf et al., 2011, Gunther et al., 2015, Kretschmer et al., 2015). Up-regulation of *TREX1*, *RNASEH2C*, and *SAMHD1* transcripts by IFN-α further supports a role for these genes in immune tolerance, by preventing the accumulation of interferogenic nucleic acids. Of note, *TREX1* was also recently associated with complex SLE in a large case-control study (Namjou et al., 2011), so it is plausible that additional monogenic lupus genes may show an association with idiopathic SLE in future larger GWAS. IFN-α also regulated the expression of *ACP5* (up-regulated) and *PRKCD* (down-regulated), both of which interact with type 1 IFN signalling components. Interestingly, functional studies in mice have shown that *PRKCD*, which encodes an isoform of protein kinase C, induces apoptosis of autoreactive B cells, and is therefore a crucial regulator of immune tolerance (Belot et al., 2013). As with the complement genes, down-regulation of *DNASE1L3* and *PRKCD* by IFN-α paralleled the genetic effects of the lupus-associated LoF variants within those genes, and may therefore contribute to the perpetuation of type 1 IFN production (Figure 4-10).

In addition to monogenic SLE, studies of various forms of syndromic lupus have advanced our understanding of the pathways that can activate type 1 IFN production. The so-called type 1 interferonopathies are a group of rare monogenic syndromes which include chilblain lupus, Aicardi-Goutières syndrome, and spondyloenchondrodysplasia. These syndromes are characterized by overproduction of IFN-α, and show a degree of clinical overlap with SLE (Crow and Manel, 2015). Studies in patients have identified LoF mutations in nucleic acid metabolism genes (including those also associated with monogenic lupus) and GoF mutations in nucleic acid sensors (*IFIH1*, *RIG-I* (*DDX58*), and *STING*), type 1 IFN inducers (*ACP5*), and IFN-inducible genes (*ISG15* and *PSMB8*). All interferonopathy genes differentially expressed by IFN-α were up-regulated, including those shared with monogenic and idiopathic SLE, which highlights their key role in driving a controlled IFN-α response in B cells from healthy donors.

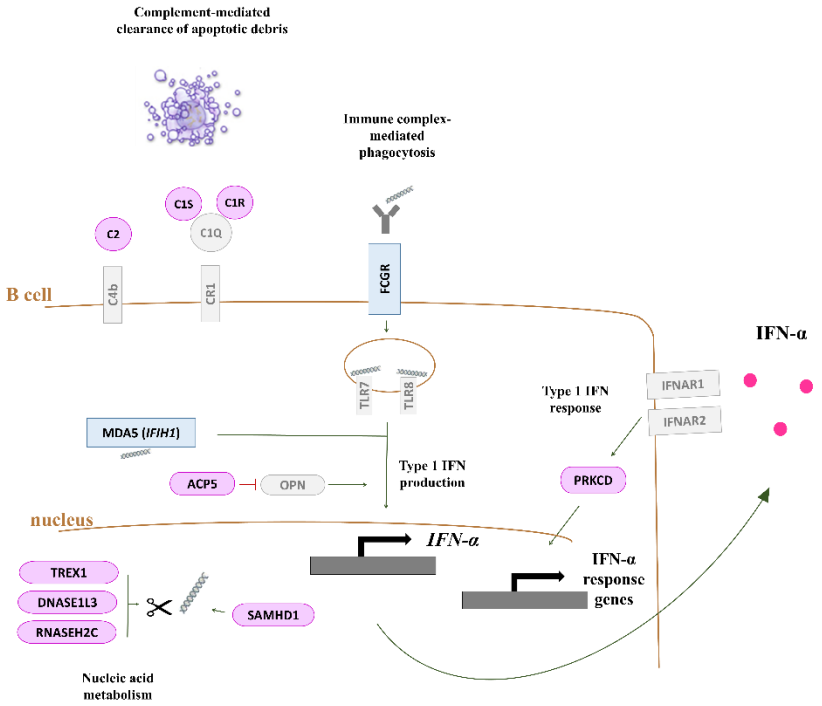


Figure 4-9 Differential expression of genes associated with monogenic lupus by IFN-α stimulation.

Monogenic lupus genes are coloured in purple. Non-differentially expressed genes relevant to the pathways are coloured in grey. The activating and inhibitory effects of lupus genes on IFN-α production are illustrated by green and red arrows, respectively.

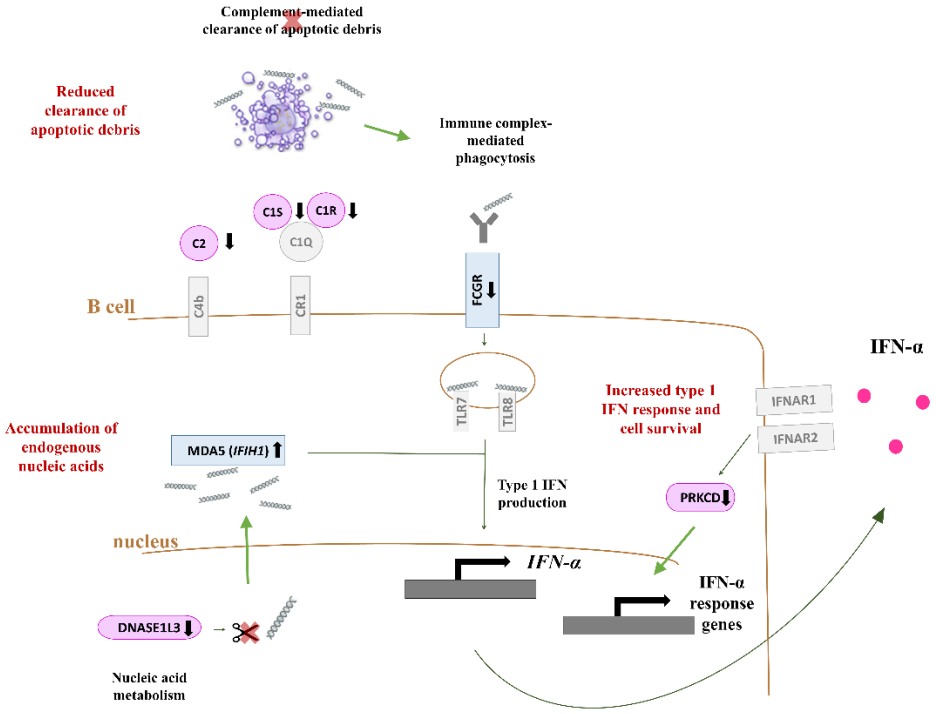


Figure 4-10 Potential consequences of IFN-α on the function of genes associated with monogenic lupus.

Associated genes are coloured in purple. Non-differentially expressed genes relevant to the pathways are coloured in grey. Black arrows indicate the direction of effect on gene expression levels following IFN-α stimulation. The resulting hypothetical increase and decrease in protein function is indicated by thick green arrows and red crosses, respectively. Changes in the expression of these genes may increase the IFN-α response through the accumulation of endogenous nucleic acids in the cytoplasm, decreased complement-mediated clearance of apoptotic debris, and increased type 1 IFN signalling.

Analysis of gene expression at the MHC showed a strong effect of IFN-α in the region, particularly on xclass I histone genes, class I and class II antigen processing and presentation genes, and class III cytokine genes. Up-regulation of HLA class I transcripts replicated previous studies in murine B cells (Braun et al., 2002, Descotes, 2004). Induction of the HLA class I in response to IFN-α is consistent with its role in anti-viral immunity, where early production of IFNs is key (Fink et al., 2006). Furthermore, IFN-α down-regulated HLA class II transcripts, including the SLE-associated *DPB1* gene. However, based on reports in the literature, the effect of IFN-α on class II molecules remains unclear, with one study showing that IFN-α up-regulated the HLA class II in buccal epithelial cells (Smith et al., 1996). In contrast, a different study reported that IFN-γ, but not IFN-α, up-regulated class II molecules in keratinocytes (Niederwieser et al., 1988).

Differential expression at the MHC also included several genes that have shown independent and non-independent associations with SLE. *MICB* (class I) and *MSH5* (class III) were the only genes regulated by IFN-α that have shown independent SNP associations with lupus (Barcellos et al., 2009, Fernando et al., 2012, Morris et al., 2012). Many reported MHC associations with SLE were no longer significant after accounting for *DRB1* alleles, due to strong LD between the loci (Clancy et al., 2010, Morris et al., 2012, Fernando et al., 2007, Barcellos et al., 2009). Interestingly, the results presented here showed that several loci within the *DRB1*03:01* extended haplotype were differentially expressed by IFN-α, including *NCR3*, *AIF1*, the TNF cytokine family genes *TNF*, *LTA*, and *LTB, C6orf10* (class II), *TRIM27* (xclass I), and *HSD17B8* (xclass II). The heat shock protein gene *HSPA1B* (class III), also differentially expressed by IFN-α, has previously shown an association with SLE independently from a *DRB1*03:01* tag SNP, but the study in question did not confirm independence from other HLA alleles (Furnrohr et al., 2010). Taken together, these data suggest that SLE-associated loci within extended MHC haplotypes may still play a biological role in the disease in the context of IFN-α activation.

It should be noted that most classical HLA class I and II genes failed QC, which highlights one of the drawbacks of using pre-designed expression arrays for highly polymorphic regions such as the MHC. qPCR analysis of *DPB1* demonstrated that this method was a reliable alternative when quantifying expression levels of this classical HLA gene.

In summary, the work described in this chapter has shown that IFN-α regulates the expression of several lupus candidate genes, including those involved in type 1 IFN signalling, BCR signalling, the classic complement pathway, nucleic acid metabolism, and HLA class I and class II antigen presentation. Differential expression of non-canonical SLE candidate genes following IFN-α stimulation suggests that they may play a previously unrecognised role in disease pathogenesis. The direction of effect of IFN-α on the expression of several non-canonical candidate genes was similar to that of functional SLE polymorphisms within those genes. In addition, the effects of IFN-α and functional risk polymorphisms also mirrored the difference in gene transcript levels between SLE patients and controls in some cases. Importantly, IFN-α decreased expression of regulators of type 1 IFN and genes involved in clearing apoptotic cell debris, ICs, and endogenous nucleic acids, all of which contribute to immune tolerance during inflammation. Based on these observations, it is possible to hypothesise that exposure to IFN-α in genetically susceptible individuals may trigger autoimmunity by altering the expression of genes involved in maintaining tolerance to self. For example, in individuals carrying an SLE risk variant that marginally decreases cellular function, IFN-α exposure could further reduce protein levels below the threshold required to sustain health, leading to the development of SLE. Altogether, these data complement ongoing functional studies of individual SLE-associated genes, and uncover potentially novel mechanisms underlying disease susceptibility.

4.6 Alternative splicing induction by IFN-α

The effects of IFN-α on gene expression so far described in this chapter have been observed in data normalized at gene level. However, analysis of gene-level data will not detect any effects on exon expression in genes where the overall transcript levels remain unchanged. Accordingly, a further analysis was performed to determine whether IFN-α affected alternative splicing in non-differentially expressed (NDE) genes. In addition, the effect of IFN-α in alternative splicing of SLE candidate genes, type 1 interferonopathy candidate genes, and genes within the xMHC region, was also investigated.

Following normalization at exon level, differential gene expression and alternative splicing between resting and IFN-α-stimulated samples was calculated. As the exon-level and gene-

level datasets were generated using different normalization methods (robust multi-array average (RMA) and winsorized mean, respectively), the list of differentially expressed genes in the two data sets was similar, but not identical (6,018 differentially expressed transcripts at gene-level compared to 7,902 transcripts at exon-level, FDR < 1%).

4.6.1 Global alternative splicing

Analysis of the global exon-level data showed that 8,708 out of 13,394 multi-exon transcripts (65%) were alternatively spliced between resting and IFN-α samples. After excluding differentially expressed transcripts (differential expression $P_{FDR} < 0.05$), 1,383 transcripts (10% of all multi-exon transcripts) were alternatively spliced.

4.6.2 Alternative splicing of SLE susceptibility genes

Out of 16 NDE SLE-associated genes, nine transcripts showed evidence of alternative splicing following IFN-α stimulation (FDR < 1%, Table 4-13, Figure 4-11). These included the canonical type 1 IFN pathway genes *IRAK1*, *TNIP1* and *IRF8*, and non-canonical genes. Interestingly, *IRF8* was the only alternative spliced gene exhibiting conflicting differential expression status between the exon-level data set (NDE, $P_{FDR} = 0.28$) and the gene-level data set (differentially expressed at an FDR < 5%, Table 4-7). Of the nine NDE genes associated with monogenic lupus, two genes showed evidence of alternative splicing, *C1QC* and *DNASE1* (FDR < 1%, Table 4-14, Figure 4-12). In contrast, none of the type 1 interferonopathy-associated genes showed evidence of alternative splicing. At the xMHC region, 14 out of 36 NDE transcripts (24%) were alternatively spliced, including the SLE-associated genes *NOTCH4* and *SKIV2L* (FDR < 1%, Appendix F, Table F2).

Table 4-13 The effect of IFN-α stimulation on alternative splicing of genes associated with idiopathic SLE.

Alternatively spliced transcript	Transcript cluster ID	P_{FDR}	Role
Strong evidence of alternative splicing (FDR < 1%)			
IRAK1	4027009	7.11×10^{-7}	TLR pathway
WDFY4	3245783	6.09×10^{-5}	Unknown
SLC15A4	3477917	4.53×10^{-4}	Peptide Transporter
FAM167A	3124388	7.24×10^{-4}	Unknown
TNIP1	2881672	1.05×10^{-3}	Type 1 IFN pathway
LRRC18	3288482	1.19×10^{-3}	Unknown
ICAM5	3820469	1.56×10^{-3}	Cell adhesion
IRF8	3672489	1.99×10^{-3}	Type 1 IFN pathway
WDFY4	3245881	2.42×10^{-3}	Unknown
Suggestive evidence of alternative splicing (FDR < 5%)			
DDA1	3824212	1.24×10^{-2}	Unknown
PRDM1	2919669	1.69×10^{-2}	IFN-β inhibition
VKORC1	3688197	1.88×10^{-2}	Vitamin K metabolism
SPRED2	2556752	2.10×10^{-2}	Growth factor signalling

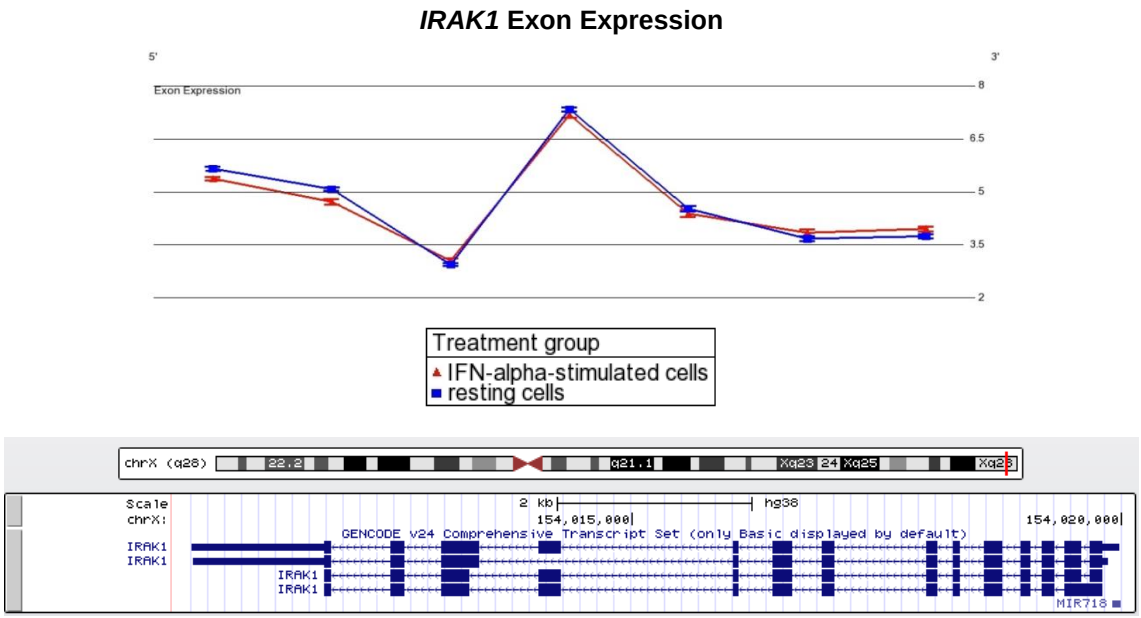


Figure 4-11 Differential exon expression in the SLE susceptibility gene *IRAK1* between resting and IFN-α-stimulated B cells.

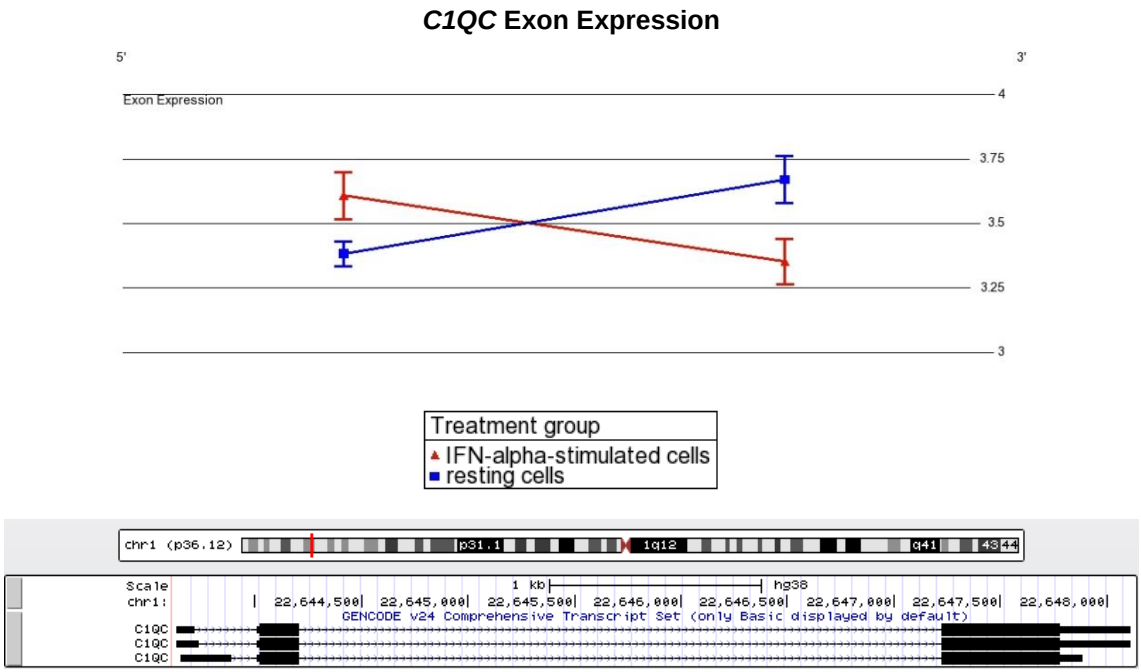
The top panel shows exon expression levels (y axis) plotted for each probe set (x axis). A UCSC screenshot with GENCODE version 24 transcript variants is shown below. Alternative splicing of *IRAK1* was significant at a P = 7.11×10^{-7} (FDR < 1%). Error bars represent the standard error of the mean.

Table 4-14 The effect of IFN-α stimulation on alternative splicing of genes associated with monogenic lupus.

Alternatively spliced transcript	Transcript cluster ID	P_{FDR}	Role
Strong evidence of alternative splicing (FDR < 1%)			
C1QC	2324873	2.09×10^{-4}	Complement
DNASE1	3646000	4.68×10^{-4}	Nucleic acid metabolism
Suggestive evidence of alternative splicing (FDR < 5%)			
PRKCD	2624291	1.10×10^{-2}	BCR-induced apoptosis
RNASEH2C	3377826	1.78×10^{-2}	Nucleic acid metabolism

BCR – B cell receptor.

A.



B.

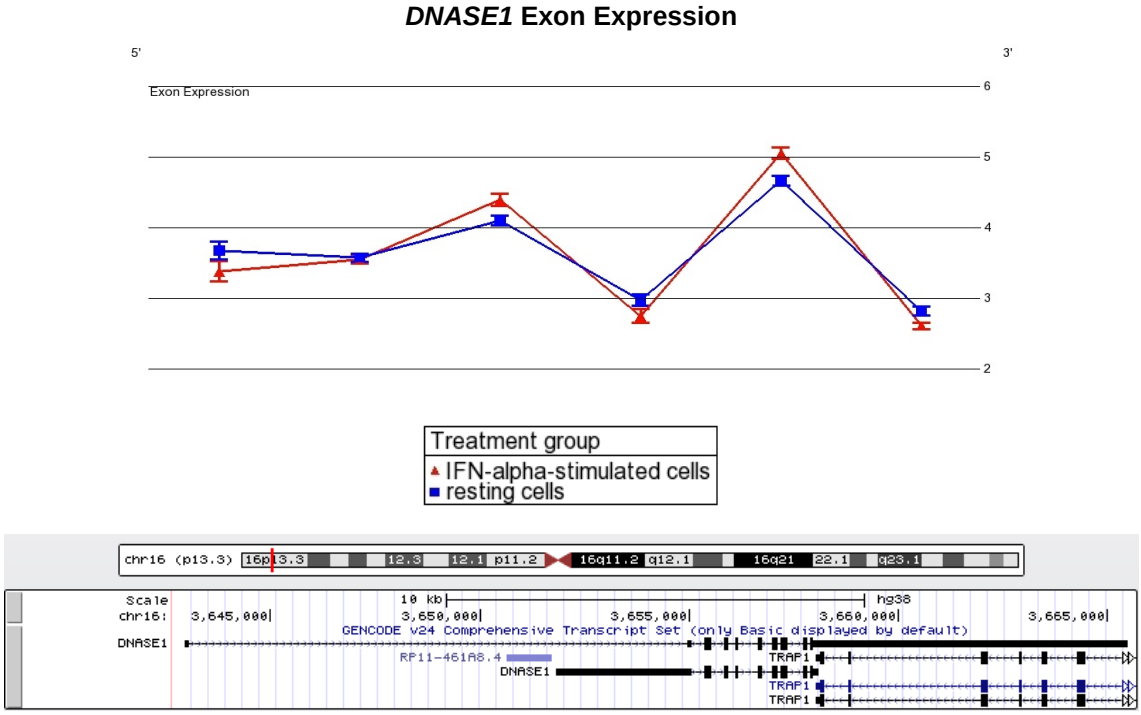


Figure 4-12 Differential exon expression in monogenic lupus susceptibility genes between resting and IFN-α-stimulated B cells.

The top panels show exon expression levels (y axis) plotted for each probe set (x axis). UCSC screenshots with GENCODE version 24 transcript variants are shown below each plot. Evidence of treatment-specific alternative splicing is shown for *C1QC* ($P = 2.09 \times 10^{-4}$, panel **A**) and *DNASE1* (4.68×10^{-4}, panel **B**). Alternative splicing events were considered significant at an FDR < 1%. Error bars represent the standard error of the mean.

4.6.3 Discussion

Genome-wide alternative splicing was shown to be significantly modified in *ex vivo* B cells following exposure to IFN-α, reflecting an additional level of molecular regulation of B cell function. Interestingly, NDE genes associated with idiopathic and monogenic SLE showed evidence of alternative splicing between resting and stimulated cells. These data implicate NDE SLE candidate genes in the type 1 IFN response, thus complementing the gene-level expression results previously presented in section 4.5. Microarray studies have previously shown that changes in alternative splicing following B cell and T cell activation can result in increased transcript diversity, or altered protein function (Grigoryev et al., 2009). The role of alternative splicing in the development of autoimmune disease has attracted increasing attention in recent years, with several studies indicating a role for alternative splicing in SLE susceptibility, as well as in disease pathogenesis (Evsyukova et al., 2010). In accordance with these observations, the data here reported implicate alternative splicing as an additional mechanism controlling the response of SLE candidate genes to IFN-α stimulation. As alternative splicing data obtained using splice ANOVA have low reproducibility, these results will require subsequent validation using different algorithms, such as splicing index (SI) or finding of isoforms using Robust Multichip Analysis (FIRMA), to detect splicing events (Srinivasan et al., 2005, Purdom et al., 2008).

Chapter 5 - eQTL analysis of MHC haplotypes in resting and IFN-α-stimulated *ex vivo* B cells

5.1 Introduction

5.1.1 Fine-mapping SLE associations at the MHC

The most consistent genetic association with systemic lupus erythematosus (SLE) in European populations arises from the human leukocyte antigen (HLA) class II allele *HLA-DRB1*03:01*, and variants in linkage disequilibrium (LD), at the major histocompatibility complex (MHC) region (Tsao, 2004, Fernando et al., 2012, International MHC and Autoimmunity Genetics Network et al., 2009). The *DRB1*03:01* allele is mainly present in two ancestral haplotypes, the *B*08:01-DRB1*03:01* haplotype (B8, common in Northern Europeans) and the *B*18:01-DRB1*03:01* haplotype (B18, common in Southern Europeans). This allele can also be found in partially-conserved B8 and B18 haplotype fragments that form due to recombination. The B8 haplotype is characterized by the specific HLA allele combination *HLA-A*01:01-Cw*07:01-B*08:01-DRB1*03:01-DQA1*05:01-DQB1*02:01*. The conservation of such extended haplotypes at the MHC is thought to have resulted from positive selection of specific HLA alleles and other variants that confer protection against infectious disease (Gluckman and Hanson, 2004, Relle and Schwarting, 2012). The B8 haplotype is associated with several autoimmune diseases in addition to SLE (Chapter 1, Table 1-2), hence its designation as the "autoimmune haplotype" (Lie and Thorsby, 2005). As the *DRB1*03:01* association with SLE has not been fine-mapped to less than ~1 megabase (MB), the causal variants underlying this signal remain to be fully elucidated (Graham et al., 2002, Fernando et al., 2012).

The identification of disease-causing variants arising from MHC haplotypes has been challenging, due to the high polymorphism of HLA genes, copy number variation, the complex LD structure in the region, and epistatic interactions between HLA and/or non-HLA alleles (Lincoln et al., 2009, Morris et al., 2012). Furthermore, the clinical heterogeneity of SLE may hamper the detection of sub-phenotype-specific associations. In order to overcome these issues, recent studies have attempted to fine-map associations at the MHC using a variety of targeted methods, such as high-density single nucleotide polymorphism (SNP) studies, trans-

ancestral analyses, meta-analyses, and sub-phenotype association studies (Barcellos et al., 2009, Fernando et al., 2012, Morris et al., 2012, Morris et al., 2014). As discussed in Chapter 3, section 3.1.1, SLE associations at the MHC have been successfully fine-mapped to a number of loci independently of *DRB1* haplotypes, including the HLA class II gene *HLA-DPB1* and the class III gene *MSH5* (Fernando et al., 2012). Despite the successful fine-mapping of a number of SLE signals arising within the MHC region, the molecular mechanisms underlying these associations have not been fully elucidated.

The current hypothesis explaining the association of *DRB1* haplotypes with SLE is that HLA class II alleles play a direct role in disease pathogenesis by affecting HLA-T cell interactions. As mentioned in previous chapters, it is possible that multiple risk factors within *DRB1* haplotypes jointly contribute to SLE susceptibility, in addition to the risk conferred by HLA alleles. This hypothesis is supported by the identification of non-HLA risk factors within *DRB1*03:01* haplotypes which may be functionally relevant to disease, such as the -308/G>A *TNF* promoter variant, or partial *C4* deletion (Abraham and Kroeger, 1999, Helmig et al., 2011, Boteva et al., 2012). Since *DRB1*03:01* haplotypes span over 1000 coding and non-coding polymorphisms, it is plausible that regulatory variants within these haplotypes may predispose to disease by altering gene expression. A number of studies have shown *cis* and *trans* regulatory effects emanating from specific HLA alleles/haplotypes (Vandiedonck et al., 2011, Fairfax et al., 2012). In addition, I have shown in Chapter 3 that SLE-associated haplotypes at the MHC harbour reproducible *cis*- and *trans*-expression quantitative trait loci (eQTLs) in B cells, lymphoblastoid cell lines (LCLs), and other blood cell populations from published studies.

5.1.2 Limitations of existing eQTL data sets

One of the weaknesses of the study conducted in Chapter 3 was the comparison of eQTLs between data sets of different cell types and cell activation states, given the context-specificity of eQTLs. Many of the early eQTL studies were performed in heterogeneous cell populations, such as whole blood and peripheral blood leukocytes; therefore, cell type-specific eQTLs may have been missed in these studies due to signal saturation from other cell populations. Other eQTL studies have used LCLs, which are cell lines derived from Epstein-Barr virus (EBV)-transformed B cells, and may therefore be partially activated. Furthermore, since B cells exhibit a high degree of cross-talk with other immune cells, analysing gene expression in *ex vivo* B

cells is more physiologically relevant than using LCLs. One of the potential issues with the *ex vivo* B cell data set of Fairfax and colleagues is the positive selection of B cells, which may have also activated the cells (Fairfax et al., 2012). Finally, the *DRB1*03:01* haplotype *cis*-eQTL to *DPB1* reported in Chapter 3 showed a heterozygote effect in the *ex vivo* B cell data set from Fairfax and colleagues; however, this may have been an artefact caused by the low number of *DRB1*03:01* homozygote samples in this data set (N = 3, Fairfax et al., 2012). For the aforementioned reasons, the eQTLs identified in B cells and LCLs described in Chapter 3 warrant further validation in B cells from healthy individuals.

One commonality between the microarray-based eQTL studies investigated in Chapter 3 was the use of Illumina platforms for gene expression quantification. Illumina arrays target the 3' untranslated region (UTR) of genes, and a limited number of annotated spliced transcripts. In contrast, the Affymetrix Human Exon 1.0 ST array (or exon array) comprises probes designed to target every exon (which together make up a probe set), thereby providing higher overall gene coverage compared to Illumina arrays. Gene-level expression data are obtained by summarising probe set data into transcript clusters. The exon array design enables a more accurate estimation of gene expression, as well as the detection of alternative splicing events that would otherwise be missed in Illumina arrays.

5.1.3 Response eQTLs

Recent eQTL studies have highlighted the importance of defining eQTLs within the context of cell stimulation, the so-called response eQTLs (reQTLs). Following exposure to a stimulus, cells exhibit more uniform global gene expression, which facilitates the discovery of *trans*-eQTLs (Fairfax et al., 2014). Furthermore, response-specific eQTLs enable the identification of gene-environment interactions that may be informative for disease. Two studies have reported an overlap between disease-associated variants and reQTLs following stimulation of immune cells with interferon (IFN)-β and IFN-γ (Lee et al., 2014, Fairfax et al., 2014). Given that IFN-α therapy is a known trigger of autoimmunity, and that type 1 IFNs play a major pathogenic role in SLE, it is plausible that SLE-associated variants may act as reQTLs in cells stimulated with IFN-α. In addition to SLE, several other autoimmune diseases have shown a type 1 IFN signature. Interestingly, of those diseases showing the signature, SLE, Sjögren's syndrome, rheumatoid arthritis, type 1 diabetes, systemic sclerosis, and idiopathic inflammatory myopathies all show

genetic association with *DRB1*03:01* haplotypes (Chapter 1, Table 1-2). Thus, the association of *DRB1*03:01* haplotypes and the type 1 IFN signature with multiple autoimmune diseases suggests that there may be a haplotype-environment interaction (H × E).

5.1.4 Aims and study design

The first aim of this study was to validate *cis*- and *trans*-eQTLs arising from SLE-associated MHC haplotypes in *ex vivo* B cells from healthy individuals, using Affymetrix exon arrays and TaqMan quantitative reverse-transcription polymerase chain reaction (qPCR). eQTLs were investigated for *DRB1*03:01* haplotypes, as well as for tag SNPs of the Spanish SLE risk *DPB1* haplotype (rs3117213), the Filipino protective *DPB1* haplotype (rs2071351), and the *MSH5* protective haplotype (rs409558), using genotype data from the TwinsUK resource.

The second aim was to investigate reQTLs in the context of IFN-α stimulation, and assess whether there are any haplotype-environment interactions between *DRB1*03:01* and IFN-α. Finally, the effects of *DRB1*03:01* haplotypes on alternative splicing were evaluated both in basal and in IFN-α-stimulated states. Study participants comprised 49 healthy European women from the TwinsUK registry. Gene expression data were obtained from negatively-selected *ex vivo* CD19+ B cells at rest and after stimulation with IFN-α for 6 hours. To increase power, analyses for resting B cells were conducted using all study samples and adjusting for IFN-α stimulation (Table 5-1).

In order to validate any eQTLs found in the CD19+ B cell data set, expression levels of relevant genes were also investigated in seven MHC-homozygous LCLs, using data obtained from exon arrays (data generated by Ms Lora Boteva and Dr Michelle Fernando, King's College London). This data set comprised two *DRB1*03:01*-homozygous LCLs (COX and QBL), and five non-*DRB1*03:01* LCLs (APD, MANN, SSTO, PGF, and DBB). The MHC haplotypes for each LCL are described in Chapter 1, Table 1-1. For all analyses, p-values were considered significant if they fell below a false discovery rate (FDR) of 5%.

Table 5-1 Sample numbers by *DRB1*, rs3117213, rs2071351, and rs409558 genotypes used for the reQTL study of resting and IFN-α-stimulated *ex vivo* B cells.

Variant	Treatment	Genotype 0 count	Genotype 1 count	Genotype 2 count	Total genotype count
DRB1*03:01	Resting	28	3	18	49
	IFN-α	21	2	9	32
	Total	49	5	27	81
rs3117213 minor allele	Resting	8	12	5	25
	IFN-α	11	5	0	16
	Total	19	17	5	41
rs2071351 minor allele	Resting	13	11	1	25
	IFN-α	11	5	0	16
	Total	24	16	1	41
rs409558 minor allele	Resting	34	4	11	49
	IFN-α	19	3	10	32
	Total	53	7	21	81

The study included a total of 49 participants. The reduced number of IFN-α-stimulated samples in the data set was due to sample loss during the experimental procedure. Genotype data for rs3117213 and rs2071351 were available for only 25 of the 49 study participants.

5.2 Quality control of exon array data

Quality control (QC) analysis of the expression data generated from twin's *ex vivo* B cells showed that the data were of good quality, as described in Chapter 4, section 4.2. For the MHC-homozygous LCL data, visualization of QC metrics using the Affymetrix Expression Console software, build 1.2.1.20 (as described in Chapter 2, section 2.4.1) revealed that all arrays had high hybridization quality, and showed normal distribution of probe intensity signals (Figure 5-1). A Spearman correlation test showed high correlation between technical replicates available for COX ($r^2 = 0.94$), QBL ($r^2 = 0.93$), PGF ($r^2 = 0.94$), and DBB ($r^2 = 0.92$).

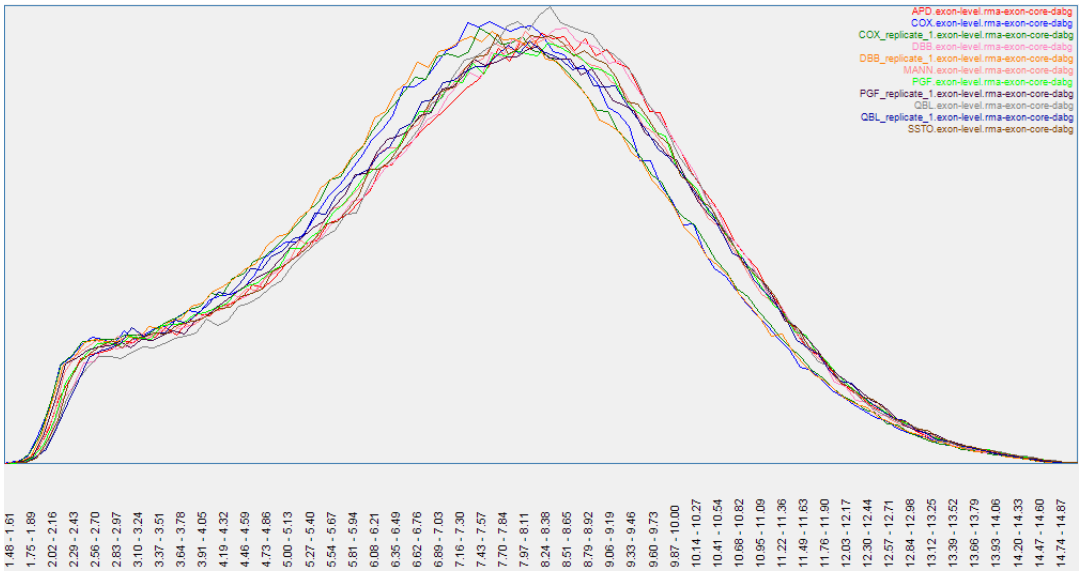


Figure 5-1 Histogram showing the normal distribution of exon array probe set intensity signals for MHC-homozygous LCL samples.

Figure generated using the Affymetrix Expression Console software, build 1.2.1.20.

Principal component analysis (PCA) of the LCL data showed that one principal component (PC) explained most of the data variability (Figure 5-2). A PCA plot revealed that batch was the biggest determinant of gene expression (Figure 5-3).

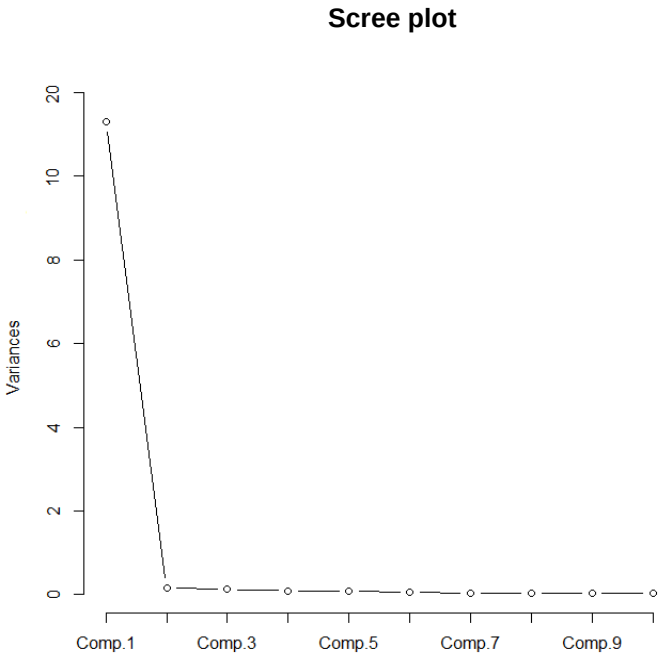


Figure 5-2 Scree plot for exon-level array data from MHC-homozygous LCLs.

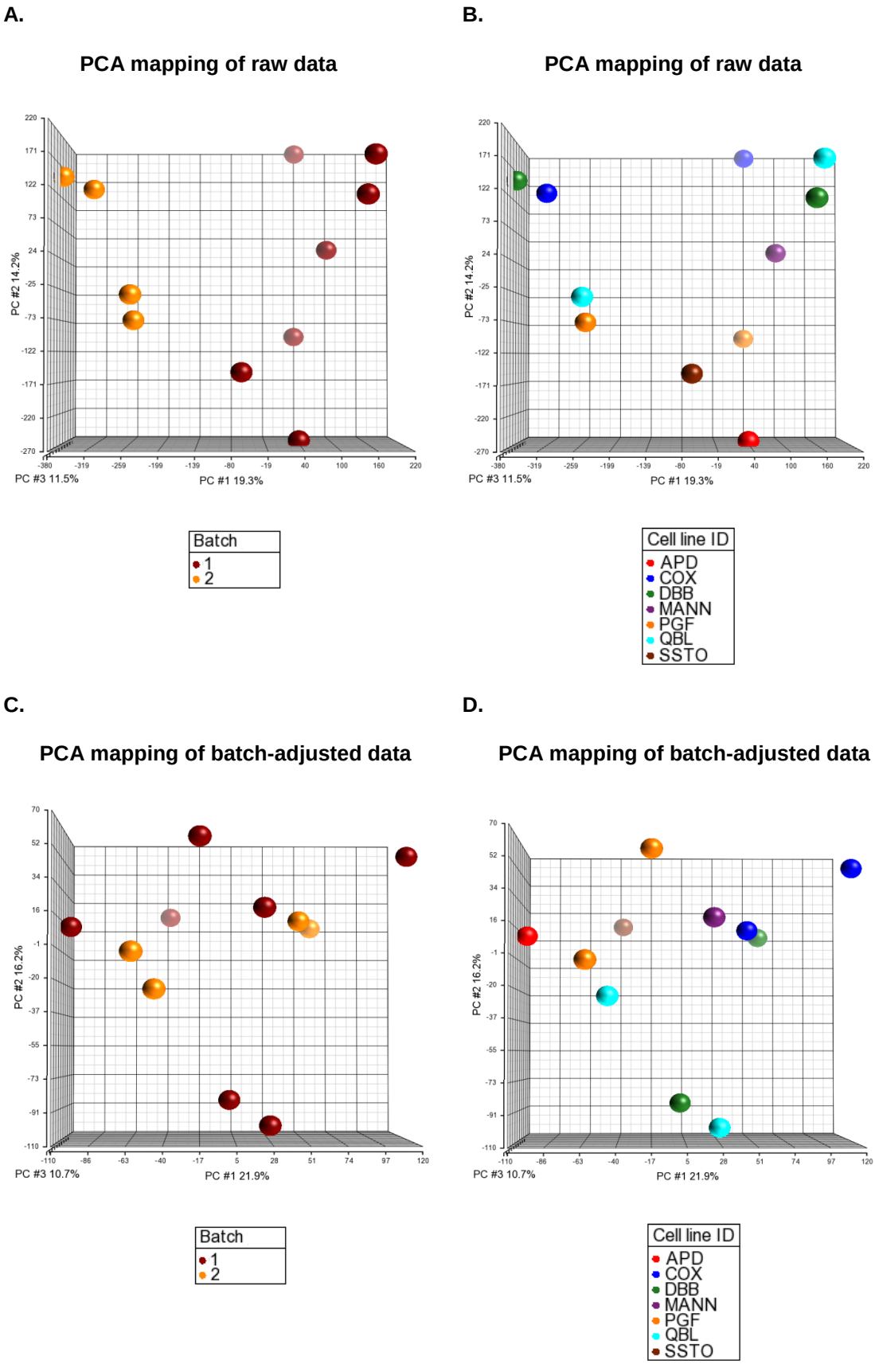


Figure 5-3 PCA plots for exon-level array data from MHC-homozygous LCLs.

Plot **A** shows raw data samples clustering by batch. Plot **B** shows raw data coloured by cell line ID. Plots **C** and **D** show the data structure following adjustment for batch effects, coloured by batch year and cell line ID, respectively.

5.3 Characterization of *HLA-DRB1*03:01* haplotypes in the study cohort

DRB1 genotypes for each study participant were initially identified using genotypic data for the *DRB1*03:01* tag SNP, rs2187668, provided by the TwinsUK resource. *DRB1* genotypes were subsequently confirmed by *DRB1* typing at four-digit resolution. In order to characterize the *DRB1*03:01* haplotypes of the study participants, the genotypes for *HLA-A, -C*, *-B*, *-DPB1*, and *-DQB1* were established using imputed HLA data, and compared to ancestral B8 and B18 haplotypes. Imputed genotypes were available for 13 *DRB1*03:01* homozygotes and one *DRB1*03:01* heterozygote. All *DRB1*03:01* haplotypes harboured the *DQB1*02:01* allele. For the *HLA-C* and *HLA-B* loci, nine *DRB1*03:01* homozygotes were *B*08-Cw*07:01* homozygotes, two were *B*08-Cw*07:01/B*18-Cw*05:01* heterozygotes, and two were *B*08-Cw*07:01*/other heterozygotes. For *HLA-A* and *DPB1*, recombinants from the ancestral B8 and B18 haplotypes were more frequent, reflecting the LD breakdown of the *DRB1*03:01* haplotype between *HLA-A* and *HLA-C*, and between *DQB1* and *DPB1*. The HLA allelic composition of *DRB1*03:01* haplotypes for each individual can be found in Appendix G. These data show that the *DRB1*03:01* haplotypes included in the study cohort are almost identical between *HLA-B* in class I and *DQB1* in class II.

5.4 *Cis*-eQTLs arising from SLE susceptibility MHC haplotypes

5.4.1 No *cis*-eQTLs arising from *HLA-DRB1*03:01* haplotypes

Due to the low number of *DRB1*03:01* heterozygotes in the exon array data set, gene expression levels were first compared between *DRB1*03:01* homozygotes and non-*DRB1*03:01* samples. A mixed-model analysis of variance (ANOVA) was performed separately for basal and IFN-α-stimulated B cells. No significant *cis*-eQTLs were found following multiple test correction in either basal or IFN-α-stimulated samples.

In Chapter 3, *DRB1*03:01* haplotypes were consistently shown to be associated in *cis* with up-regulation of *DPB1* and *PSMB9* in publicly-available data sets. Validation of these *cis*-eQTLs was therefore attempted in *ex vivo* basal and IFN-α-stimulated B cells from TwinsUK. *DPB1*

expression could not be investigated using the exon array data set, as all probe sets for this gene failed QC. Differential expression of *DPB1* was therefore investigated by TaqMan qPCR, using complementary deoxyribonucleic acid (cDNA) from study samples. The pre-designed TaqMan assay for *DPB1* does not target regions harbouring common SNPs. The number of remaining cDNA samples was smaller than the original cohort used for the exon array experiment (N = 32 non-*DRB1*03:01* samples, N = 3 *DRB1*03:01* heterozygotes, and N = 13 *DRB1*03:01* homozygotes). *DRB1*03:01* homozygote and non-*DRB1*03:01* haplotypes did not show significant differential expression of *DPB1* in resting cells (P = 0.22, fold change (FC) = 1.14, ANOVA). However, there is a clear trend towards up-regulation of *DPB1* in *DRB1*03:01* homozygotes, as shown in Figure 5-4, with a similar fold change to that reported in Chapter 3, section 3.3 (FC ~ 1.3). Since there was only one remaining *DRB1*03:01* homozygote sample from IFN-α-stimulated cells in sufficient amount for qPCR, the analysis could not be carried out in this group. For *PSMB9*, neither the exon array data nor the qPCR analysis suggested differential expression in basal or IFN-α-stimulated cells (P > 0.98).

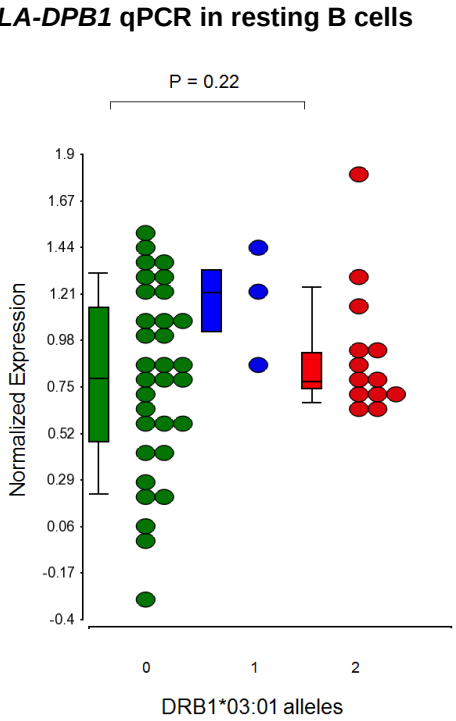


Figure 5-4 qPCR analysis of *HLA-DPB1* expression in *HLA-DRB1*03:01* haplotypes in resting *ex vivo* B cells.

The plot shows a trend for up-regulation of *DPB1* in *DRB1*03:01* homozygotes and heterozygotes, compared to non-*DRB1*03:01* samples (FDR > 5%). Expression levels were normalized against the control genes *UBB*, *TBP*, and *OAZ1*. The box plots show median, upper, and lower quartiles of expression levels for each genotype group. Whiskers show the maximum and minimum non-outlying data points.

5.4.2 No *cis*-eQTLs arising from *HLA-DPB1* and *MSH5* haplotypes

Cis-eQTL analyses were performed for the SLE-associated *DPB1* Spanish (rs3117213) and Filipino (rs2071351) SNPs using qPCR data, in an attempt to validate *cis*-associations with *DPB1* expression described in Chapter 3. Out of the 48 available qPCR samples, only 41 had genotype data for rs3117213 and rs2071351 (Table 5-1). *Cis*-eQTL analysis showed that there were no significant eQTLs for either rs3117213 or rs2071351; however, the eQTL plots revealed a clear trend for up-regulation of *DPB1* in rs3117213 heterozygotes, compared to homozygotes for the ancestral allele, in both resting and IFN-α-stimulated cells (P > 0.29, FC = 1.05, Figure 5-5). For rs2071351, this trend was only observed in IFN-α-stimulated cells (P = 0.45, FC = 1.08, Figure 5-6). Similarly, analysis of the exon array data showed that there were no significant *cis*-eQTLs with the SLE-protective *MSH5* SNP, rs409558, in either resting or IFN-α-stimulated B cells.

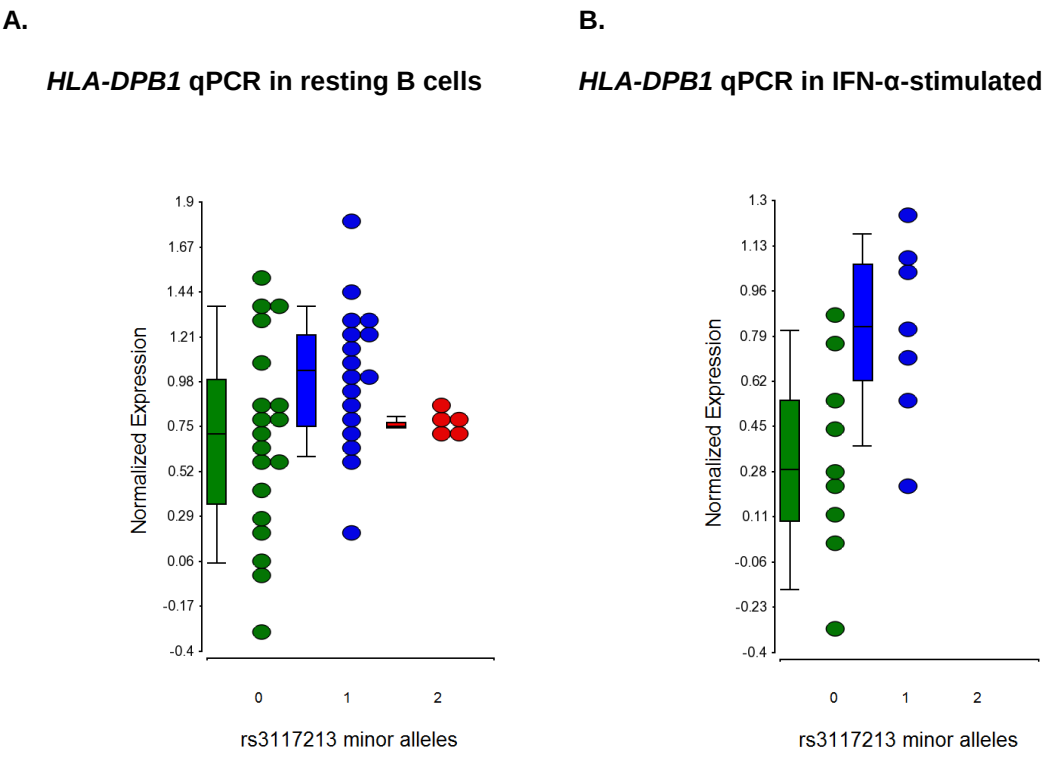


Figure 5-5 qPCR analysis of *HLA-DPB1* expression for rs3117213 genotypes in resting and IFN-α-stimulated *ex vivo* B cells.

The plots show a trend for up-regulation of *HLA-DPB1* in heterozygotes, compared to homozygotes for the ancestral allele, in B cells at rest (**A**) and following stimulation with IFN-α (**B**, FDR > 5%). Expression levels were normalized against the control genes *UBB*, *TBP*, and *OAZ1*. The box plots show median, upper, and lower quartiles of expression levels for each genotype group. Whiskers show the maximum and minimum non-outlying data points.

A.

***HLA-DPB1* qPCR in resting B cells**

B.

***HLA-DPB1* qPCR in IFN-α-stimulated B cells**

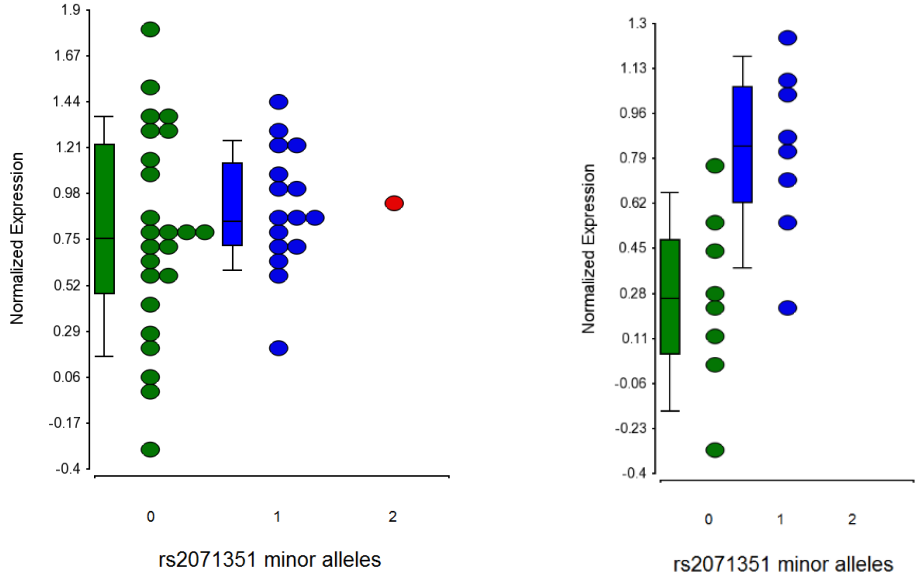


Figure 5-6 qPCR analysis of *HLA-DPB1* expression for rs2071351 genotypes in resting and IFN-α-stimulated *ex vivo* B cells.

A. No evidence of differential expression between rs2071351 genotypes was found in resting B cells. **B.** The plot shows a trend for up-regulation of *HLA-DPB1* in rs2071351 heterozygotes, compared to homozygotes for the ancestral allele, in IFN-α-stimulated B cells (FDR > 5%). Expression levels were normalized against the control genes *UBB*, *TBP*, and *OAZ1*. The box plots show median, upper, and lower quartiles of expression levels for each genotype group. Whiskers show the maximum and minimum non-outlying data points.

5.5 *Trans*-eQTLs arising from *HLA-DRB1*03:01* haplotypes

5.5.1 Association between *HLA-DRB1*03:01* and *BTN3A2* expression

Trans-eQTL analysis for the *DRB1*03:01* haplotype in basal cells revealed two significantly differentially expressed transcript clusters, both of which mapped to the *BTN3A2* gene (Figure 5-7). These two transcript clusters, TC_2899298 and TC_2899333, showed under-expression in *DRB1*03:01* homozygote samples, compared to non-*DRB1*03:01* samples ($P < 2.34 \times 10^{-8}$, FC < -2.00, Figure 5-8, Table 5-2). In IFN-α-stimulated cells, only TC_2899298 showed significant down-regulation following multiple test correction ($P = 2.63 \times 10^{-6}$, FC = -2.10, Table 5-3). Differential expression of *BTN3A2* was then analysed using a linear regression for comparison between *DRB1*03:01* homozygotes, heterozygotes, and non-*DRB1*03:01* samples. Differential

expression of *BTN3A2* was significant prior to multiple test correction in basal cells ($P = 9.31 \times 10^{-3}$ for TC_2899298, $P = 2.05 \times 10^{-2}$ for TC_2899333, FDR > 5%), but not in IFN-α-stimulated cells ($P > 0.05$).

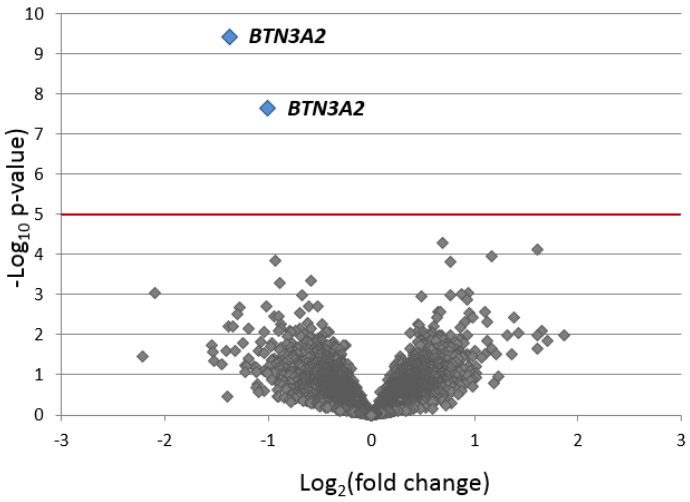


Figure 5-7 Volcano plot showing *trans*-eQTLs for the *HLA-DRB1*03:01* haplotype in resting *ex vivo* B cells.

Trans-associations are plotted as $-\log_{10}$(p-values) against $\log_2$(fold change). Two *BTN3A2* transcript clusters from the Affymetrix Human Exon 1.0 ST array showed significant differential expression between *DRB1*03:01* homozygotes and non-*DRB1*03:01* samples. The red line indicates the significance threshold at an FDR of 5%.

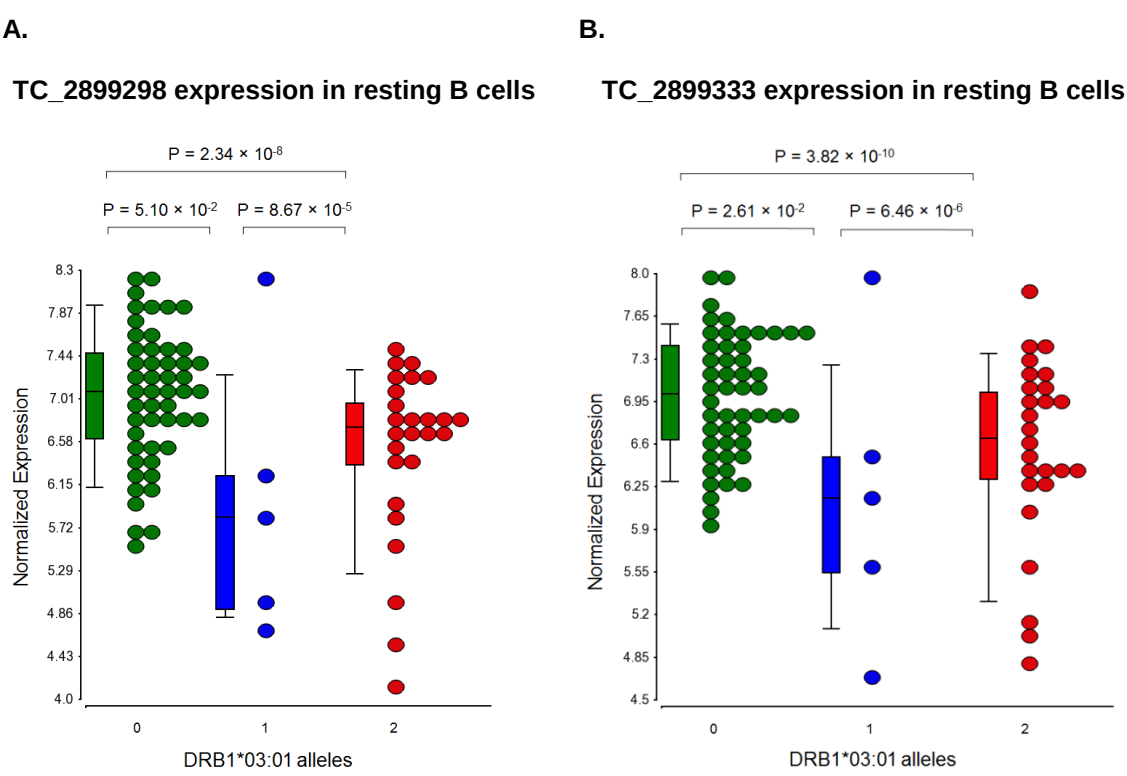


Figure 5-8 The *HLA-DRB1*03:01* haplotype is an eQTL for *BTN3A2* in resting *ex vivo* B cells.

The Affymetrix Human Exon 1.0 ST array transcript clusters for *BTN3A2*, TC_2899298 (**A**) and TC_2899333 (**B**), were significantly down-regulated in *DRB1*03:01* homozygotes, compared to non-*DRB1*03:01* samples (FDR < 5%). In contrast, there was no significant differential expression between non-*DRB1*03:01* and *DRB1*03:01* heterozygotes, or between *DRB1*03:01* heterozygotes and homozygotes (FDR > 5%). The box plots show median, upper, and lower quartiles of expression levels for each genotype group. Whiskers show the maximum and minimum non-outlying data points.

Table 5-2 ANOVA results for *BTN3A2* expression between *HLA-DRB1*03:01* and non-*DRB1*03:01* haplotypes in resting *ex vivo* B cells.

***DRB1* genotype comparisons**	**TC_2899298**			**TC_2899333**		
	Fold change	**P-value**	**P_{FDR}**	**Fold change**	**P-value**	**P_{FDR}**
****03:01* homozygotes *vs.* non-**03:01* samples**	-2.60	2.34×10^{-8}	1.90×10^{-4}	-2.00	3.82×10^{-10}	6.19×10^{-6}
****03:01* heterozygotes *vs.* non-**03:01* samples**	-1.67	5.10×10^{-2}	0.99	-1.52	2.61×10^{-2}	0.99
****03:01* homozygotes *vs.* **03:01* heterozygotes**	-1.59	8.67×10^{-5}	0.35	-1.34	6.46×10^{-6}	5.23×10^{-2}

BTN3A2 expression levels were quantified using the Affymetrix Exon 1.0 ST array, and summarised as two transcript clusters, TC_2899298 and TC_2899333. The significance threshold was established at an FDR of 5%.

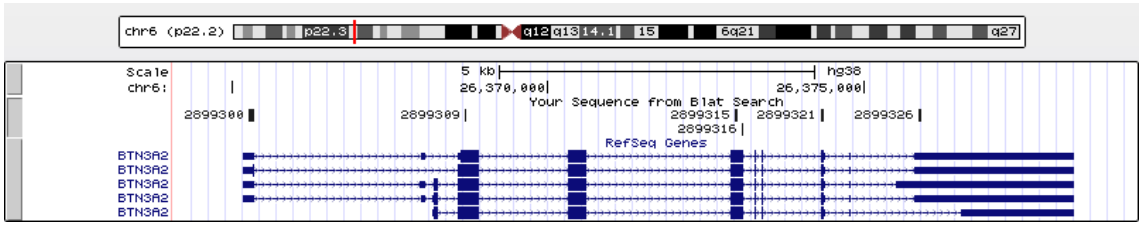
Table 5-3 ANOVA results for *BTN3A2* expression between *HLA-DRB1*03:01* and non-*DRB1*03:01* haplotypes in IFN-α-stimulated *ex vivo* B cells.

***DRB1* Genotype Comparisons**	**TC_2899298**			**TC_2899333**		
	Fold change	**P-value**	**P_{FDR}**	**Fold change**	**P-value**	**P_{FDR}**
****03:01* homozygotes *vs.* non-**03:01* samples**	-2.10	2.63×10^{-6}	4.26×10^{-2}	-1.96	1.20×10^{-5}	0.10
****03:01* heterozygotes *vs.* non-**03:01* samples**	-1.11	0.70	0.80	-1.85	7.75×10^{-2}	0.57
****03:01* homozygotes *vs.* **03:01* heterozygotes**	-1.54	1.80×10^{-4}	0.48	-1.74	6.36×10^{-3}	0.86

BTN3A2 expression levels were quantified using the Affymetrix Exon 1.0 ST array, and summarised as two transcription clusters, TC_2899298 and TC_2899333. The significance threshold was established at an FDR of 5%.

The two *BTN3A2* transcript clusters target different regions of the gene: TC_2899298 targets the 5' UTR through to the 3' UTR, whereas TC_2899333 exclusively targets the 3' UTR (Figure 5-9).

A.



B.

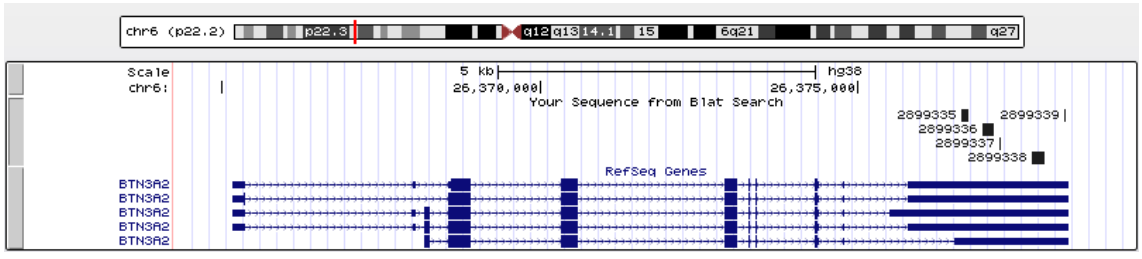


Figure 5-9 UCSC Genome Browser screenshot showing the *BTN3A2* target sites for the Affymetrix Human Exon 1.0 ST array probe sets included in the *ex vivo* B cell analysis.

Probe set sequences for each *BTN3A2* transcript cluster were aligned to the genome using the BLAT tool, and are shown above *BTN3A2* transcripts from RefSeq. **A**. The TC_2899298 transcript cluster comprises six probe sets, which target the 5' UTR, exons 4, 6, 9, and the 3' UTR, and detect both full-length and truncated transcripts. **B**. TC_2899333 comprises five probe sets, which target the 3' UTR, and detect all full-length *BTN3A2* transcripts.

The *BTN3A2* gene is located in the extended MHC (xMHC) class I region, within a cluster of butyrophilin 3 (BTN3) genes: *BTN3A1*, *BTN3A2*, and *BTN3A3* (Figure 5-14). Clustal alignment of the cDNA sequences for all three BTN3 genes revealed high sequence homology between them (89-94%, Table 5-4, Appendix H).

Table 5-4 Clustal alignment scores for the coding sequences of the paralogue butyrophilin genes *BTN3A1*, *BTN3A2*, and *BTN3A3*.

Gene	*BTN3A1*	*BTN3A2*	*BTN3A3*
BTN3A1	-	88.86%	89.69%
BTN3A2	88.86%	-	94.13%
BTN3A3	89.69%	94.13%	-

Alignment scores are displayed as percentages.

As a result of the high sequence homology between the three BTN3 genes, quantification of *BTN3A2* expression by conventional Illumina arrays and RNA sequencing (RNA-seq) data sets is not reliable. The Illumina probes annotated for *BTN3A2* cross-hybridize to *BTN3A1* and *BTN3A3*, and therefore fail QC in eQTL studies using Illumina platforms (including LCL, *ex vivo* B cell, monocyte, and whole blood data sets investigated in Chapter 3). Additionally, due to the technical challenges of processing RNA-seq data from highly homologous genes, the eQTL for *BTN3A2* could also not be validated in publicly-available RNA-seq data sets. To confirm that the *BTN3A2* eQTL identified using the exon array was not an artefact, validation of the results by TaqMan qPCR was attempted, using cDNA from the study samples. A pre-designed TaqMan qPCR assay specific for *BTN3A2* transcripts was obtained from Life Technologies. The assay target region did not map to common SNP-harbouring regions, and spanned a sequence of low homology to *BTN3A1* and *BTN3A3* (Figure 5-10). Down-regulation of *BTN3A2* in *DRB1*03:01* homozygotes compared to non-*DRB1*03:01* samples was validated by qPCR in basal B cells ($P = 2.06 \times 10^{-3}$, FC = -2.88, FDR < 5%). In IFN-α-stimulated B cells, this eQTL trended towards significance at a similar fold change ($P = 6.46 \times 10^{-2}$, FC = -2.22, FDR > 5%, Figure 5-11).

A.

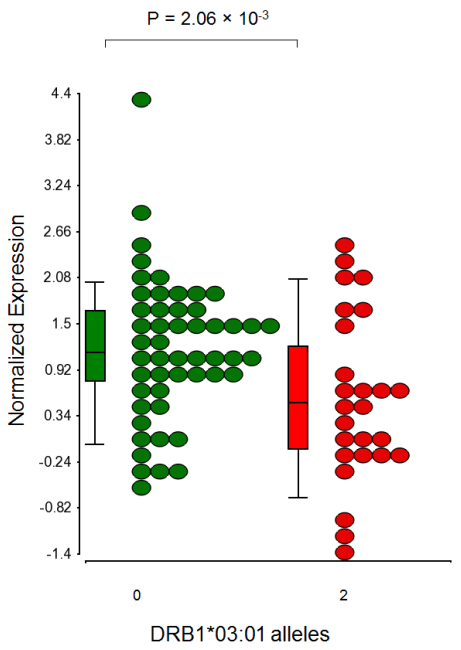
B.

```
BTN3A1      GATGGAGAAGTATCCAGTATGCATCTCGGGGAGAGAGACATTCAGCCTATAATGAATGGAAAAAGGCC
BTN3A2      AGAGGAAAAAAATCCAGTACTTGACTCGTGGAGAGGAGTCTTCGTCCGATACCAATAAGTCAGCCTGA
BTN3A3      AGTGGAGGAAAATCCAGTACATGGCTCGTGGAGAGAAGTCTTTGGCCTATCATGAATGGAAAATGGCC
            ****   ***  *  ******** **** ******     **   ** **    *   *  *
```

Figure 5-10 Location of the *BTN3A2* region targeted by the TaqMan qPCR assay Hs00389328_m1.

A. The Hs00389328_m1 assay (shown in green) targets a region within exons 8 and 9 of the *BTN3A2* gene, represented below (image obtained from Life Technologies). The exact assay target sequence is not specified by Life Technologies. **B**. Clustal cDNA sequence alignment showing the comparison of exons 8 and 9 of *BTN3A2* with *BTN3A1* and *BTN3A3*. *BTN3A2* exon 8 is highlighted in yellow and exon 9 is highlighted in cyan. The asterisks represent sequence homology between *BTN3A1*, *BTN3A2*, and *BTN3A3*.

A.



B.

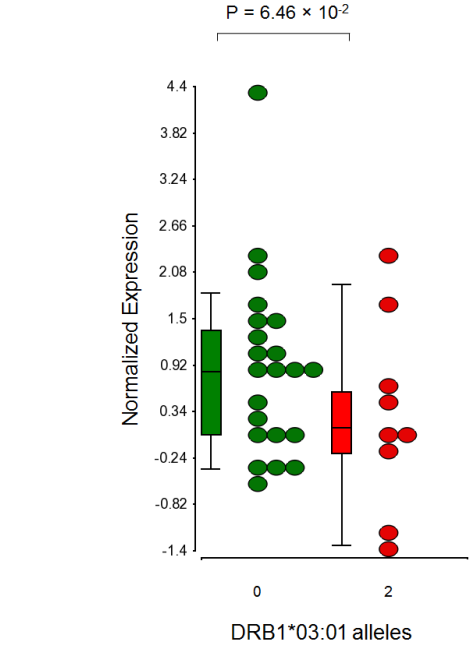


Figure 5-11 qPCR analysis validated the down-regulation of *BTN3A2* in *DRB1*03:01* homozygotes compared to non-*DRB1*03:01* samples from *ex vivo* B cells.

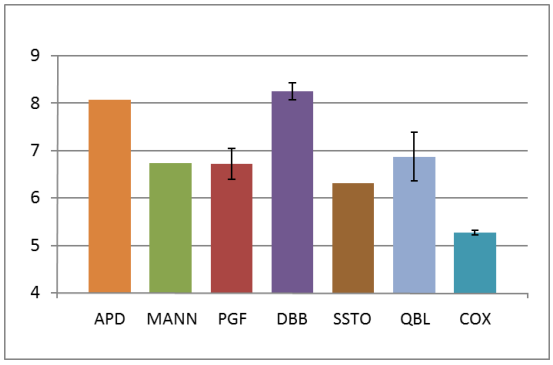
BTN3A2 down-regulation in resting cells was significant at an FDR < 5% (**A**), but not in IFN-α-stimulated cells (FDR > 5%, **B**). Expression levels were normalized against the control genes *UBB*, *TBP*, and *OAZ1*. The box plots show median, upper, and lower quartiles of expression levels for each genotype group. Whiskers show the maximum and minimum non-outlying data points.

Validation of the eQTL for *BTN3A2* in MHC-homozygous LCLs at the protein-level was then attempted. First, *BTN3A2* messenger ribonucleic acid (mRNA) levels were assessed in seven LCLs using data obtained from exon arrays. This data set comprised two *DRB1*03:01*-homozygous LCLs (COX, N = 2, and QBL, N = 2), and five non-DRB1*03:01 LCLs (APD, N = 1; MANN, N= 1; SSTO, N = 1; PGF, N = 2; and DBB, N = 2), including technical replicates. Differential expression of *BTN3A2* between *DRB1*03:01* and non-*DRB1*03:01* LCLs was not significant (P > 0.13, FC = -2.53, ANOVA). Interestingly, down-regulation of *BTN3A2* in COX (*B*08-DRB1*03:01*) compared to other LCLs showed a trend towards significance (TC_2899298, $P = 8.37 \times 10^{-2}$, FC = -3.76, FDR > 5%, Figure 5-12).

Given the effect sizes found for the *DRB1*03:01 trans*-eQTL to *BTN3A2* in exon array and qPCR experiments, the translation of this eQTL to protein level was investigated in MHC-homozygous LCLs. A Western blot was performed to detect protein expression in two *DRB1*03:01*-homozygous LCLs (COX and QBL) and three non-*DRB1*03:01* LCLs (APD, MANN, and PGF). The *BTN3A1* and *BTN3A3* gene products (BT3.1 and BT3.3, respectively) were also examined for comparison with the *BTN3A2* gene product (BT3.2). Antibodies for each of the butyrophilin proteins were specific for the C terminus regions. The amino acid antigen for the BT3.2 antibody encompassed the same exons targeted by the TaqMan qPCR assay (exons 8 and 9, Appendix I). Protein levels were quantified by normalization against β-actin, and levels were compared between LCLs using a Student's *t*-test. BT3.2 showed evidence of down-regulation in COX and QBL, compared to non-*DRB1*03:01* LCLs ($P = 3.88 \times 10^{-2}$, Figure 5-13). In contrast, BT3.1 and BT3.3 were not differentially expressed in LCLs (P > 0.40).

A.

TC_2899298 expression in resting LCLs



B.

TC_2899333 expression in resting LCLs

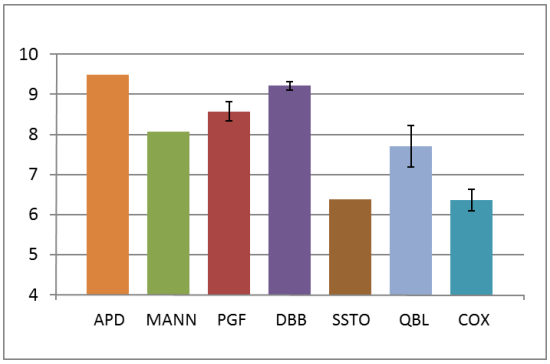


Figure 5-12 *BTN3A2* expression in seven MHC-homozygous LCLs at rest, quantified by Affymetrix Human Exon 1.0 ST arrays.

Expression levels for the TC_2899298 transcript cluster are shown in **A**, and expression levels for TC_2899333 are shown in **B**. *BTN3A2* was not significantly differentially expressed between *DRB1*03:01*-homozygous LCLs (COX and QBL) and non-*DRB1*03:01* LCLs (APD, MANN, PGF, DBB, and SSTO, $P > 0.13$). TC_2899298 showed a trend towards being down-regulated in COX (*B*08-DRB1*03:01*) compared to other LCLs ($P = 8.37 \times 10^{-2}$, non-significant). Error bars represent the standard error of the mean for samples where duplicate data were available.

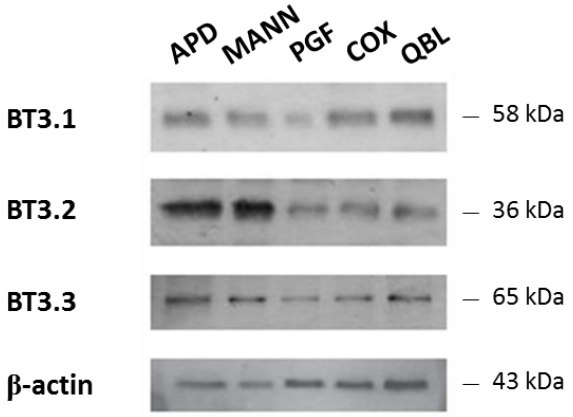


Figure 5-13 BT3.1, BT3.2, BT3.3, and β-actin protein expression detected by Western blotting in five MHC-homozygous LCLs.

Western blot analysis showed evidence of down-regulation of BT3.2 in two *DRB1*03:01* LCLs (COX and QBL), compared to three non-*DRB1*03:01* LCLs (APD, MANN, and PGF, $P = 3.88 \times 10^{-2}$). In contrast, BT3.1 and BT3.3 protein levels were not significantly altered between *DRB1*03:01* and non-*DRB1*03:01* LCLs ($P = 0.68$ and $P = 0.40$, respectively). kDa – kilodalton.

Interestingly, the *BTN3A2* gene in xMHC class I is approximately 6 MB from the *DRB1* gene in class II (Figure 5-14). To investigate whether the *DRB1*03:01* eQTL with *BTN3A2* is caused by LD with a *cis*-eQTL at *BTN3A2*, a *cis*-eQTL analysis was performed for *BTN3A2* on resting B cells. Out of 47 tested SNPs, none were significant eQTLs after multiple test correction. Thirty-six SNPs, however, correlated with decreased *BTN3A2* levels at p-values of less than 0.05 prior to multiple test correction (best SNP P = 1.85 × 10^{-4}, FDR > 5%, FC = -1.36). All SNPs are in low LD with the *DRB1*03:01* proxy SNP rs2187668 (r^2 < 0.09, TwinsUK cohort). Thus, it is unclear whether the *DRB1*03:01* eQTL with *BTN3A2* is a true *trans*-eQTL, or whether it is tagging a *cis*-eQTL that was not genotyped in the cohort.

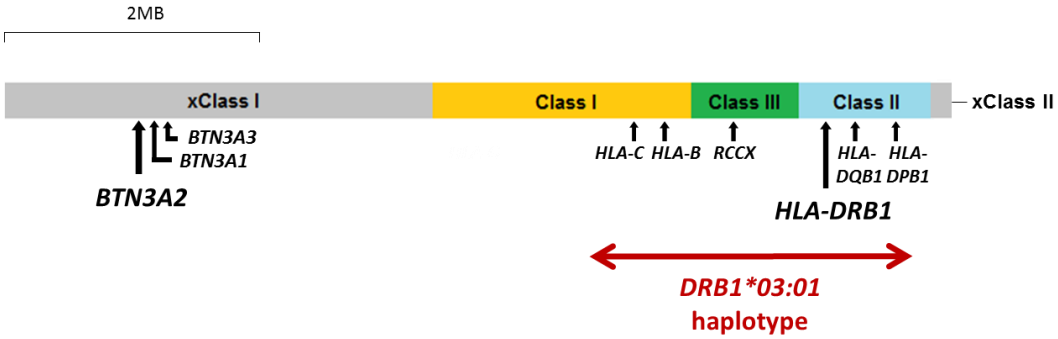


Figure 5-14 Location of the BTN3 and *DRB1* genes within the xMHC region on chromosome 6.

5.5.2 No evidence of interaction between *HLA-DRB1*03:01* haplotypes and IFN-α

An interaction ANOVA was performed to test whether *DRB1*03:01* homozygotes and non-*DRB1*03:01* individuals responded differently to IFN-α stimulation at the mRNA level. The analysis showed no evidence of interaction at gene level between the two factors (P_{FDR} > 0.46).

5.6 Effect of *HLA-DRB1*03:01* haplotypes on alternative splicing

The effect of harbouring *DRB1*03:01* haplotypes on alternative splicing was assessed, as described in Chapter 4, section 4.5. Partek's splice ANOVA in resting B cell samples did not show evidence of alternative splicing between *DRB1*03:01* homozygotes and non-*DRB1*03:01* samples (using a significance threshold of FDR < 5%). In contrast, analysis of IFN-α-stimulated

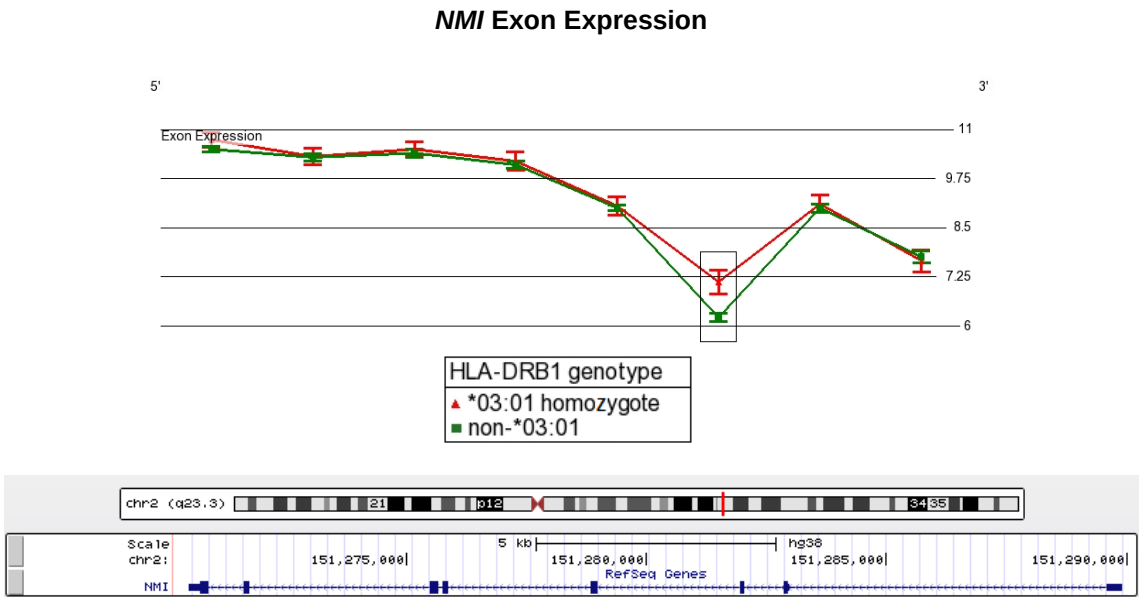
cells revealed that 30 out of 13,543 multi-exon transcript clusters were alternatively spliced between *DRB1*03:01* and non-*DRB1*03:01* haplotypes at significant levels (FDR < 5%, Table 5-5). Interestingly, a great portion of the alternatively spliced genes are type 1 IFN-inducible genes, such as *STAT1* and *OAS2*. Ingenuity pathway analysis (IPA) of alternatively spliced genes confirmed enrichment of IFN signalling ($P = 1.55 \times 10^{-7}$) and bacteria and virus pattern recognition receptor pathways ($P = 1.48 \times 10^{-8}$). Additionally, the IFN-α gene *IFNA2* was predicted as the top upstream regulator ($P = 4.54 \times 10^{-26}$). There was no significant enrichment of idiopathic or monogenic SLE-associated genes in the data set ($P > 0.19$). Alternative splicing of two IFN-inducible transcripts is illustrated in Figure 5-15.

Table 5-5 Alternatively spliced genes between *HLA-DRB1*03:01* and non-*DRB1*03:01* haplotypes in IFN-α-stimulated *ex vivo* B cells.

Alternatively spliced transcript	Transcript cluster ID	P_{FDR}	Role
STAT1	2592268	3.11×10^{-8}	Type 1 IFN pathway
DDX58 (RIG-I)	3203086	3.85×10^{-6}	Type 1 IFN pathway
IFIT1	3257246	3.97×10^{-6}	Type 1 IFN response
SEPT9	3735847	3.73×10^{-5}	Cell cycle
HERC6	2735362	3.73×10^{-5}	Type 1 IFN response
SBF2	3362468	4.64×10^{-4}	Unknown
ISG15	4053534	4.83×10^{-4}	Type 1 IFN response
MX2	3922037	6.05×10^{-4}	Type 1 IFN response
PLSCR1	2699726	6.05×10^{-4}	Type 1 IFN response
TARS	2805786	6.05×10^{-4}	tRNA synthesis
OAS2	3432514	7.42×10^{-4}	Type 1 IFN response
EIF2AK2	2548402	1.14×10^{-3}	Type 1 IFN response
TDRD7	3181193	1.21×10^{-3}	Type 1 IFN response
TLR7	3969081	1.31×10^{-3}	Type 1 IFN pathway
TIMM10	3373946	1.34×10^{-3}	Mitochondrial transport
SP100	2531377	5.46×10^{-3}	Type 1 IFN response
IFIH1	2584207	5.54×10^{-3}	Type 1 IFN pathway
MX1	3922100	5.54×10^{-3}	Type 1 IFN response
CD4	3402786	6.85×10^{-3}	T cell activation
PNPT1	2553970	8.72×10^{-3}	Type 1 IFN response
GBP4	2421995	1.10×10^{-2}	Type 1 IFN response
NMI	2580955	1.29×10^{-2}	Type 1 IFN response
OAS3	3432467	1.43×10^{-2}	Type 1 IFN response
RNF138	3783749	1.43×10^{-2}	Unknown
INPP4B	2787459	1.50×10^{-2}	Glycerophospholipid signalling
MFAP1	3621692	1.70×10^{-2}	Unknown
LCK	2328841	1.82×10^{-2}	T cell development
ZFP36L1	3569754	1.84×10^{-2}	Unknown
LOC100862671	3656418	2.01×10^{-2}	Unknown
CD6	3332663	4.19×10^{-2}	T cell activation

Alternative splicing events were considered significant at an FDR < 5%.

A.



B.

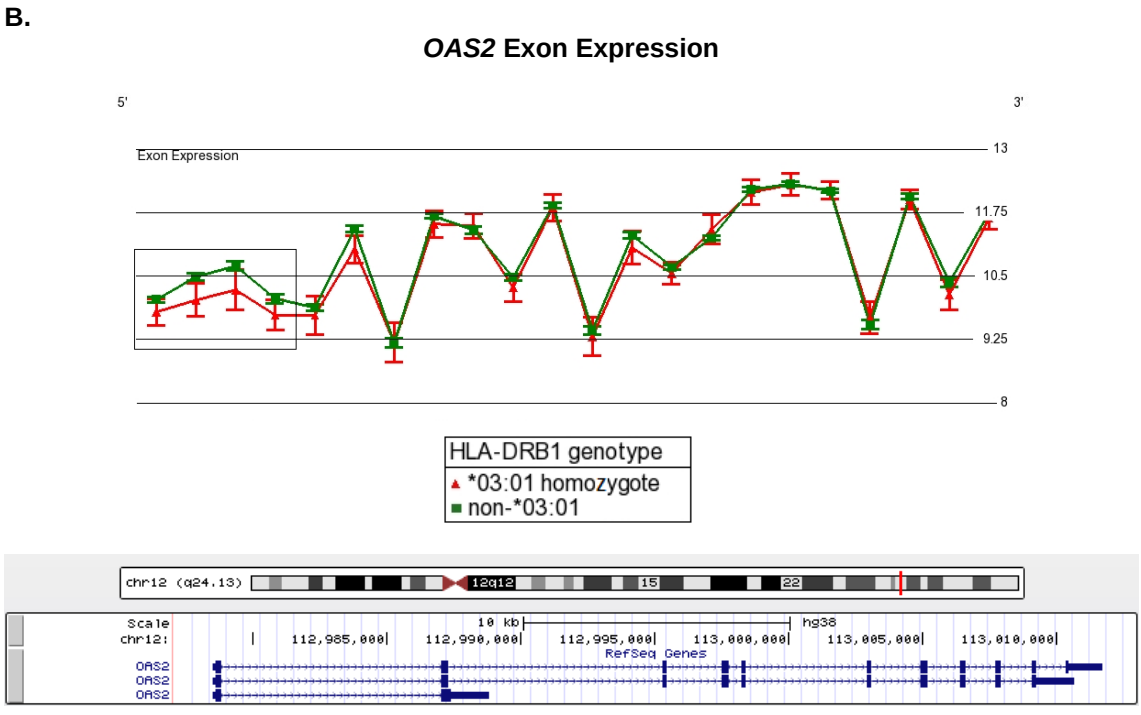


Figure 5-15 Differential exon expression in two IFN-inducible genes between *HLA-DRB1*03:01* homozygotes and non-*DRB1*03:01* samples.

The top panels show exon expression levels (y axis) plotted for each probe set (x axis). UCSC screenshots with RefSeq-annotated transcript variants are shown below each plot. **A**. Evidence of haplotype-specific cassette exon usage (boxed region) in *NMI* ($P = 2.11 \times 10^{-5}$, FDR < 5%). **B**. Evidence of haplotype-specific expression of a truncated splice transcript variant (boxed region) in *OAS2* ($P = 6.05 \times 10^{-7}$, FDR < 5%). Error bars represent the standard error of the mean.

5.7 Discussion

In this chapter, gene expression patterns for SLE-risk *DRB1*03:01* haplotypes were characterized in *ex vivo* B cells from healthy women, using the Affymetrix Human Exon 1.0 ST array. *Cis*-eQTL analysis did not validate differential expression of *DPB1*, or any other gene, in *DRB1*03:01* haplotypes at a significant level, despite a clear trend for up-regulation of *DPB1* in this cohort. *Trans*-eQTL analysis of basal and IFN-α-stimulated B cells showed significant down-regulation of *BTN3A2* in *B*08-DRB1*03:01* haplotypes, at both mRNA- and protein-level. There was no evidence of haplotype-environment interaction between *DRB1*03:01* haplotypes and IFN-α stimulation at gene level. Interestingly, *DRB1*03:01*-homozygous samples showed alternative splicing of several IFN-inducible genes following IFN-α-stimulation, but not in resting cells, suggesting a haplotype-environment interaction at exon-level.

None of the previously reported *cis*-eQTLs for *DRB1*03:01* haplotypes were validated in this study, despite a clear trend towards up-regulation of *DPB1* in resting B cells, at a similar fold change as that found in other data sets in Chapter 3, section 3.3. Similarly, *cis*-eQTLs reported for the rs3117213 and rs2071351 *DPB1* variants (section 3.4), and the rs409558 *MSH5* variant (section 3.2), were not validated. Again, a trend towards up-regulation of *DPB1* was observed for rs3117213 and rs2071351 in resting and IFN-α-stimulated B cells. These negative results suggest that, despite the fact that there is a larger number of *DRB1*03:01* homozygotes in this study than in any of the previously studied data sets, there was insufficient power to detect all true eQTLs for *DRB1*03:01*, rs3117213, rs2071351, or rs409558 haplotypes.

Association of *DRB1*03:01* haplotypes with down-regulation of *BTN3A2* was the only significant *trans*-eQTL detected in this study. This eQTL was found in both resting and IFN-α-stimulated B cells using the exon array, and was validated by qPCR in resting B cells. The trend for down-regulation of *BTN3A2* observed in qPCR data from IFN-α-stimulated B cells supports the robustness of this eQTL. Exon array quantification of *BTN3A2* expression in MHC-homozygous LCLs showed a trend towards decreased *BTN3A2* expression levels in COX (*B*08-DRB1*03:01* haplotype), compared to six other LCLs. Western blot analysis validated down-regulation of *BTN3A2* at the protein-level in both COX and QBL (*B*18-DRB1*03:01* haplotype), compared to non-*DRB1*03:01* LCLs. Since most *ex vivo* B cell samples were *B*08-DRB1*03:01*

homozygotes, these results strongly support a correlation between this haplotype and *BTN3A2* expression. Given the location of *BTN3A2* in the xMHC class I region, and the high LD encompassing the region, it is possible that the reported eQTL may be tagging a *cis*-eQTL at *BTN3A2*. Due to the low *BTN3A2* genotype coverage available for this cohort, *cis*-eQTLs within *BTN3A2* could not be identified; therefore, the mechanism underlying the eQTL for the *DRB1*03:01* haplotype remains unclear.

Expression of *BTN3A2* could not be reliably detected in previous eQTL studies, due to the high sequence homology between the BTN3 genes. However, eQTL analysis of two Illumina array studies in blood cells, and one RNA-seq study in LCLs, showed *trans*-association between the *DRB1*03:01* European tag SNP, rs2187668, and *BTN3A2* expression at an FDR < 5%, despite being potentially confounded by signals from *BTN3A1* and *BTN3A3* (Zeller et al., 2010, Westra et al., 2013, Lappalainen et al., 2013). The validity of the results here presented is supported by the specific targeting of *BTN3A2* by the exon array probe sets, the qPCR assay, and the Western blot antibody. Furthermore, no differential expression of *BTN3A1* or *BTN3A3* was detected in the exon array and Western blot analyses. Future work would be required to replicate this eQTL in *ex vivo* B cells from an independent cohort, ideally by qPCR analysis.

As alterations in mRNA levels do not always translate to protein level due to translational repression mechanisms, validation of eQTLs at the protein level has been challenging, particularly for *trans*-eQTLs (Parts et al., 2014, Albert and Kruglyak, 2015). Therefore, the validation of the *BTN3A2* eQTL at protein-level here described is of great interest, and puts forward BT3.2 as a plausible target protein for treating autoimmune disease.

BTN3A2 encodes one of three isoforms of CD277 (BT3), an immunoglobulin-like cell surface receptor belonging to the B7 family of co-stimulatory molecules. There are a total of seven butyrophilin genes in humans, which are positioned in tandem across the xMHC class I region: *BTN1A1* (single-copy), BTN2 genes (*BTN2A1*, *BTN2A2*, and *BTN2A3*), and BTN3 genes (*BTN3A1*, *BTN3A2*, and *BTN3A3*); and one butyrophilin-like (BTNL) gene located in the MHC class II (*BTNL2*, Rhodes et al., 2001). Despite having undergone gene duplication, butyrophilin gene numbers are stable in humans (Salomonsen et al., 2014). Butyrophilins have attracted strong interest in recent years due to their role as T cell-regulating molecules (Arnett et al.,

2009, Abeler-Dorner et al., 2012). CD277 molecules have been shown to act as co-inhibitors of T cell proliferation and T helper type 1 (Th1) cytokine production when expressed by antigen-presenting cells (Cubillos-Ruiz and Conejo-Garcia, 2011). Other reports have shown that these molecules play a role in the recognition of stress responses by γδ T cells (Harly et al., 2012, Vavassori et al., 2013); however, the individual function of each butyrophilin member has not yet been fully elucidated. The CD277 isoforms BT3.1 (*BTN3A1*) and BT3.3 (*BTN3A3*) contain a cytoplasmic domain (B30.2), whereas BT3.2 (*BTN3A2*) lacks this domain, and is thought to act as a decoy receptor (Rhodes et al., 2001). BT3.2 is expressed in lymphocytes and natural killer (NK) cells, and is the only CD277 isoform that can inhibit NK cell-induced production of cytokines (Messal et al., 2011). *BTN3A2* has not previously shown genetic association with SLE; however, variants within various butyrophilin and butyrophilin-like genes have been linked with several autoimmune and inflammatory diseases, including lupus (*BTNL2*), anti-citrullinated protein antibody (ACPA)-positive rheumatoid arthritis (*BTN2A2* and *BTN3A1*), type 1 diabetes (*BTN3A2*), myositis (*BTNL2*), sarcoidosis (*BTNL2*), and inflammatory bowel disease (*BTNL2,* Orozco et al., 2005, Orozco et al., 2011, Viken et al., 2009, Scott et al., 2011, Spagnolo et al., 2007, Pathan et al., 2009), and also with chronic kidney disease (*BTN2A1*, Yoshida et al., 2011). Given the role of co-inhibitory molecules in the prevention of autoimmunity (Watanabe and Nakajima, 2012), the down-regulation of *BTN3A2* in SLE-risk *DRB1*03:01* haplotypes reported here is of functional interest. Hypothetically, down-regulation of BT3.2 at the cell surface could result in decreased T cell inhibition and, consequently, lower the threshold for autoimmune responses in the presence of self-antigens.

Haplotype-environment interaction analysis showed that *DRB1*03:01* and non-*DRB1*03:01* haplotypes did not respond differently to IFN-α stimulation at gene level, suggesting a lack of interaction between haplotype and cell stimulus. Conversely, there may be insufficient power to detect an interaction. Interestingly, *DRB1*03:01* homozygote samples showed evidence of alternative splicing following IFN-α stimulation, which involved several IFN-inducible genes. These results demonstrate that *DRB1*03:01*-homozygous and non-*DRB1*03:01* individuals respond differently to IFN-α, i.e., there is a haplotype-environment interaction at exon level, but not at gene level. Alternatively, these results could be explained by the increased expression levels of IFN-inducible genes following IFN-α stimulation, which may have provided greater sensitivity to detect changes in gene splicing. As most existing alternative splicing analysis

pipelines (including splice ANOVA) do not provide high true-positive detection rates, these results require validation using different alternative splicing calculation methods, such as the splicing index (SI) and finding of isoforms using Robust Multichip Analysis (FIRMA) algorithms (Srinivasan et al., 2005, Purdom et al., 2008).

The use of *ex vivo* B cells from healthy female individuals, both at rest and after IFN-α stimulation, provided a disease-relevant context for the exploration of eQTLs in SLE-risk haplotypes, given the interaction between genetic, hormonal, and environmental factors in disease pathogenesis. However, one of the major drawback of this study is its limited power, which may have prevented the identification of highly significant signals. The limitations highlighted in section 3.6 of Chapter 3 regarding the use of SLE signals derived from non-European cohorts, and the use of expression data sets generated from European populations to functionally annotate these SLE signals, also apply to the study described in this chapter. Furthermore, given the strong effect of gender on gene expression, the use of females in this study may have also limited the replication of eQTLs found in publicly-available data bases, since most of these studies were not adjusted for gender (Chapter 3, Table 3-1 and Table 3-3). An additional limitation of the study relates to the potential effects of age and medication, for which data were not available.

Despite these limitations, the use of the Affymetrix exon expression array proved to be effective for the identification of a novel eQTL in *ex vivo* B cells. The wide gene coverage provided by exon arrays enabled quantification of gene expression for *BTN3A2* and its paralogous genes, which could not be reliably detected using Illumina arrays or RNA-seq. However, one major caveat to the use of microarrays, including the exon array, is the high polymorphism and copy number variation at the MHC, which challenge accurate detection of gene expression in the region. Since conventional microarrays are designed to target the reference human genome MHC sequence (obtained from the PGF cell line), expression levels of most *HLA* genes, including *DPB1*, could not be quantified in this study. RNA-seq provides an alternative method for the detection of genome-wide gene expression; however, the method faces similar challenges to those involved in the use of microarrays when resolving expression data at the MHC. In addition, the processing of RNA-seq data is computationally intensive, and short read lengths generated by RNA-seq hamper the resolution of copy number variable regions, such as

the MHC. The challenges of reliably detecting gene expression at the MHC may be overcome with the use of targeted qPCR assays (as shown in this study), custom microarrays, or fluorescence-activated cell sorting (FACS). The detection of *DPB1* expression by qPCR described in this chapter demonstrates that this method is a reliable alternative to measure the expression levels of this classical HLA gene.

One final limitation of the study design is that the comparison between *DRB1*03:01* and non-*DRB1*03:01* haplotypes may hinder the identification of true eQTLs, due to the length and variability of MHC haplotypes. The majority of non-*DRB1*03:01* individuals are heterozygous for classical HLA genes. As HLA genes can have large numbers of existing alleles in the population, the non-*DRB1*03:01* haplotypes included in this study will be highly variable. Most of the *DRB1*03:01* haplotypes from the study cohort were almost identical between *HLA-C* in class I and *DQB1* in class II, and demonstrated LD breakdown between *HLA-A* and *HLA-C*, and between *DQB1* and *DPB1*, which is typical of *DRB1*03:01* haplotypes in the British population (Bishof et al., 1993). Therefore, the lack of homogeneity between the haplotypes tested in the study may affect the detection of eQTL signals. Despite these caveats, the data generated for this study may be used in future eQTL meta-analysis studies of B cells, given the large number of *DRB1*03:01*-homozygous samples in the cohort.

In conclusion, this study supports a regulatory role for the *DRB1*03:01* haplotype in B cells, and implicated *BTN3A2* as a novel candidate gene underlying the association of the *DRB1*03:01* haplotype with autoimmunity. In addition, this study implicates *DPB1* as a highly likely *cis*-eQTL, further supporting the results presented in Chapter 3. The study also highlights the importance of specifically targeting paralogous genes, such as the BTN3 genes, as well as the need for large sample sizes, in order to detect and corroborate eQTL signals at the MHC. Investigating the effects of disease-associated haplotypes on gene regulation under relevant cellular contexts, such as the study here presented, will further our understanding of the molecular mechanisms underlying susceptibility to complex diseases.

Chapter 6 - Conclusions

The work described in this thesis investigates the regulatory effects of systemic lupus erythematosus (SLE) susceptibility factors in B cells, namely risk and protective haplotypes, and interferon (IFN)-α stimulation. Firstly, expression quantitative trait locus (eQTL) analyses were conducted, using published and newly-generated data sets, in order to investigate whether SLE-associated variants within the major histocompatibility complex (MHC) region correlate with altered gene expression in healthy individuals of European ancestry. Secondly, gene regulation following IFN-α stimulation was characterized in B cells, with particular focus on differential expression of disease susceptibility genes.

6.1 *Cis*- and *trans*-eQTLs arising from MHC haplotypes

Chapters 3 and 5 describe the identification of several *cis*-eQTLs for the SLE-associated haplotypes *DRB1*03:01*, *DRB1*15:01*, rs3117213 (*DPB1*), rs2071351 (*DPB1*), and rs409558 (*MSH5*), in lymphoblastoid cell lines (LCLs) and *ex vivo* B cells. Reproducible *cis*-associations identified in different blood cell types include *DOM3Z*, *C6orf48*, and *HSPA1B* for rs409558, and *HLA-DPB1* and *PSMB9* for *DRB1*03:01* haplotypes. Furthermore, functional annotation of the *MSH5* protective haplotype highlighted the best functional single nucleotide polymorphism (SNP) candidates, rs409558 and rs4711279. These results implicate a regulatory role for disease-associated variants within MHC haplotypes, and pinpoint possible functionally-relevant genes implicated in the pathogenesis of SLE. In particular, *cis*-associations with increased *DPB1* expression in individuals harbouring *DRB1*03:01* haplotypes, as well as rs3117213 and rs2071351 minor alleles, suggest an important role for this classical human leukocyte antigen (HLA) class II gene in susceptibility to SLE. It can be hypothesised that the functional consequence of these disease-associated *DPB1* eQTLs in B cells may lead to increased HLA class II antigen presentation to CD4+ T cells, thus potentially clarifying the role of *DPB1* in SLE, and other autoimmune diseases.

6.1.1 The effect of *DRB1*03:01* haplotypes on gene regulation

One other major finding is the *trans*-association of *B*08-DRB1*03:01* haplotypes with down-regulation of *BTN3A2* in resting and IFN-α-stimulated B cells, at both mRNA- and protein-level. Since mRNA levels do not always translate to protein, validation of this eQTL at protein-level in LCLs is of great functional interest, particularly with regard to the role of the *BTN3A2* gene product, BT3.2, in T cell inhibition. Down-regulation of BT3.2-mediated co-inhibitory signals in *DRB1*03:01*-homozygous B cells could result in decreased inhibition of T cell activation and/or co-stimulation, potentially leading to a state of immune hyper-responsiveness in the presence of self-antigens. This scenario is also consistent with the known role of co-inhibitory molecules in the prevention of autoimmunity. Furthermore, this is one of a small number of studies that have successfully translated *trans*-eQTLs to protein-level in primary cells. These results implicate *BTN3A2* as a novel candidate gene underlying the association of the *DRB1*03:01* haplotypes with autoimmunity, and propose BT3.2 as a potential therapeutic target for disease.

The investigation of gene regulation within the extended *B*08-DRB1*03:01* haplotype in SLE-relevant cellular conditions provided additional insights into the functional effects of this important disease susceptibility haplotype. Although no response eQTLs were identified at gene level following stimulation of B cells with IFN-α, the data suggested a possible alteration of transcript splicing patterns in response to IFN-α stimulation. On an unrelated note, stimulation of B cells with IFN-α altered the expression levels of several MHC genes, including those that harbour SLE-risk variants present in *DRB1*03:01* haplotypes, such as *DPB1*, *TNF*, *LTA*, and *LTB*, as described in Chapter 4; this further supports a functional role for these genes in the pathogenesis of SLE.

The presence of eQTLs emanating from the MHC region, particularly from disease-associated *DPB1* and *DRB1* haplotypes, supports a role for HLA haplotypes in regulation of gene expression, in addition to their probable effect on HLA amino acid variation**.** Thus, these data support the hypothesis described in Chapter 1, which states that the large effect size of genetic associations with disease observed at the MHC may be a result of the combined effect of HLA allelic variation and regulatory variants**.**

6.1.2 Study design

The eQTL results here presented also underline the importance of study design, particularly the study power and the choice of gene expression platforms, in achieving reproducibility. The eQTL study conducted in *ex vivo* B cells using the Affymetrix Human Exon 1.0 ST Array (described in Chapter 5) highlighted the need for large sample sizes in order to corroborate the *cis*- and *trans*-eQTLs identified in published studies (Chapter 3), as exemplified by the lack of significance for the *cis*-eQTLs at *DPB1*. This work also highlights the importance of specifically targeting homologous genes, such as the butyrophilin 3 (BTN3) genes, in eQTL studies, which had not previously been possible using Illumina arrays or RNA sequencing (RNA-seq) technologies. The exon array proved to be an effective platform for the specific detection of the BTN3 homologs, which allowed the identification of a novel *trans*-association with *BTN3A2*. However, a major downside of using microarrays for MHC studies is the heterogeneity of MHC haplotypes in the population, which results in impaired or inaccurate detection of gene expression. Quantitative reverse-transcription polymerase chain reaction (qPCR) was the only method that reliably detected the expression of specific MHC genes which failed quality control in published array data sets (such as *BTN3A2*), or in the exon array data (such as the polymorphic *DPB1* gene). qPCR may therefore be a preferable method for the detection of highly polymorphic and/or highly homologous genes, as opposed to commercial or custom expression microarrays and RNA-seq. Nonetheless, the exon array study described in Chapter 5 successfully reproduced the eQTL to *BTN3A2* using three different methodologies, i.e. exon array, qPCR, and Western blotting, thus supporting the use of a combination of different methodologies to validate novel eQTLs.

The work presented in Chapter 3 and Chapter 5 also reiterates the importance of context-specificity, particularly cell type-specificity, in the detection and validation of *trans*-eQTLs. The novel regulatory effects arising from SLE-associated haplotypes described in the exon array study were identified under hormonal, cellular, and environmental conditions relevant to the disease, i.e. the selection of female individuals, the use of *ex vivo* B cells, and stimulation with IFN-α. Thus, the identification of eQTLs for SLE susceptibility haplotypes in disease-relevant contexts here reported provides new insights into the candidate genes and molecular mechanisms that are causally involved in SLE.

6.1.3 Future work

One of the caveats of this study was that the SLE-associated SNPs here investigated are not necessarily the causal SNPs, which may partly explain why this study failed to identify reproducible eQTLs. Furthermore, corroboration of the genetic associations in *MSH5* and *DPB1* with Spanish and Filipino SLE is still required in larger studies. These associations were shown to be in low linkage disequilibrium (LD) with each other and with classical HLA alleles in Chapter 3. Thus, the identification of the causal alleles for these associations with SLE may be facilitated by the LD breakdown in different populations. To ensure that these SLE signals are truly independent from HLA alleles, the correlation between the associated *MSH5* and *DPB1* SNPs and HLA amino acid variation should also be investigated.

The *MSH5* and *DPB1* eQTLs reported in this thesis require validation in independent studies using populations of similar ancestry to those used in the original association study, due to the variation of the eQTL landscape across different ancestries. Any of the eQTLs here reported may be the result of weak or moderate LD between the tested variant and the peak expression SNP (eSNP) for the target transcripts. Therefore, the LD and genetic distance between the SNP of interest and peak eSNP of target transcripts need to be assessed. Similarly, *BTN3A2 cis*-eQTLs in mild LD with *DRB1*03:01* haplotypes are likely to explain the *trans*-association with *BTN3A2*, and should therefore be investigated in densely-genotyped, well-powered eQTL studies in healthy individuals and/or SLE patients, using exon arrays, exon-level RNA-seq, or qPCR.

Given the relatively large number of *DRB1*03:01*-homozygous samples included in the exon array data set, the expression data generated for this eQTL study might be incorporated in future eQTL meta-analyses conducted in B cells, in order to increase the power to detect significant eQTLs for *DRB1*03:01* haplotypes. Future validation of *cis*- and *trans*-eQTLs at the MHC, particularly to *DPB1* and *BTN3A2*, may benefit from being executed using qPCR, in order to bypass the aforementioned issues related to highly homologous and/or polymorphic genes. The validation of alternative splicing data would ideally be performed in a well-powered qPCR study, due to the low true-positive detection rates of splicing algorithms. Lastly, the functional role of the variants underlying these eQTLs might be investigated by looking for causal rare variants in LD with peak eSNPs, or variants that alter transcription factor binding sites.

Investigation of mRNA and protein expression levels of *cis*- and *trans*-associated genes in B cells from patients may also further our understanding of the molecular mechanisms through which eQTLs affect disease susceptibility.

6.2 The effect of IFN-α on the regulation of disease-associated genes

Chapter 4 explored the effects of IFN-α stimulation on gene expression in B cells, in order to elucidate the mechanism by which type 1 IFNs might confer susceptibility to autoimmunity. Acute IFN-α stimulation strongly affected transcriptome modulation, including the triggering of differential expression of several B cell markers, which provided some insight into B cell phenotypic alterations. Interestingly, IFN-α altered the expression of susceptibility genes for the autoimmune diseases SLE, rheumatoid arthritis, inflammatory bowel disease, and multiple sclerosis, but not for non-autoimmune diseases. Since relatively few studies have addressed the role of type 1 IFN in inflammation and tissue destruction in autoimmune diseases other than SLE, these data provide a platform for further investigation of the effects of IFN-α in the pathogenesis of these diseases.

IFN-α stimulation altered the expression of several SLE candidate genes involved in type 1 IFN signalling, B cell receptor signalling, the classic complement pathway, nucleic acid metabolism, and HLA class I and class II antigen presentation. Interestingly, the direction of effect of IFN-α on the expression of several genes associated with idiopathic and monogenic lupus mirrored the genetic consequences of risk polymorphisms reported in functional studies. In some cases, the effects of IFN-α also mirrored the difference in transcript levels between SLE patients and controls. Differential expression of SLE candidate genes that had never been functionally associated with the IFN-α response (and B cell receptor signalling) suggests that these genes play a previously unrecognised role in disease pathogenesis through activation of the type 1 IFN pathway. IFN-α decreased the expression of regulators of type 1 IFNs, as well as genes involved in clearing apoptotic cell debris, immune complexes, and endogenous nucleic acids (such as *ETS1*, *ITGAM*, *FCGR* genes, *DNASE1L3*, and complement component genes), all of which are important in the maintenance of immune tolerance during inflammation. Thus,

exposure of genetically susceptible individuals to IFN-α may trigger autoimmunity by altering the expression of genes involved in maintaining tolerance to self. Furthermore, SLE candidate genes were also shown to undergo alternative splicing in B cells following exposure to IFN-α, reflecting an additional level of molecular regulation of B cell function. This implicates alternative splicing in the control of the response of SLE candidate genes to IFN-α stimulation, in addition to gene expression, thus complementing the aforementioned gene-level results. Altogether, these data contribute to the ongoing functional investigation of SLE-associated genes in disease, by elucidating potential mechanisms through which IFN-α contributes to the initiation and perpetuation of autoimmunity.

6.2.1 Future work

In order to determine whether the alteration in the expression of SLE candidate genes by IFN-α translates to protein level, these expression changes need to be validated by fluorescence-activated cell sorting (FACS) or Western blotting. Investigating protein levels in *ex vivo* B cells from SLE patients and healthy controls, and whether these alterations mirror those induced by IFN-α stimulation, would also be of interest and provide further insights into the role of SLE candidate genes and IFN-α in disease pathogenesis.

In summary, the eQTL results presented in this thesis implicate a regulatory role for SLE-associated variants within MHC haplotypes, and identify potential functionally-relevant genes in the pathogenesis of SLE, namely *BTN3A2* and *DPB1*. This work also highlights the importance of study design in order to achieve reproducibility in eQTL studies. Additionally, the differential expression of SLE candidate genes following activation of the type 1 IFN pathway suggests a previously unrecognised role for these genes in disease. Taken together, these results have helped to further elucidate the mechanisms underlying genetic and environmental associations with SLE and other autoimmune diseases.

Appendix A

Table A1 Genotype count for seven SLE-associated SNPs in the *ex vivo* B cell data set from Fairfax et al. (2012).

SNP	Location on chromosome 6 (gene)	Minor allele	Genotype 0 count	Genotype 1 count	Genotype 2 count	Minor allele frequency
rs2293861	31819103 (*MSH5*)	T	193	77	11	17.62%
rs2187668	32713862 (*HLA-DQA1*)	T	262	56	3	9.66%
rs3131379	31829012 (*MSH5*)	A	228	47	3	9.53%
rs3135391	32518965 (*HLA-DRA*)	A	216	56	9	13.17%
rs3117213	33172583 (*HLA-DPB1*)	T	145	120	16	27.04%
rs2071351	33151908 (*HLA-DPB1*)	G	182	93	6	19.45%
rs2071349	33151498 (*HLA-DPB1*)	G	210	66	5	13.52%

All genomic coordinates are from the NCBI Build 36 genomic assembly.

Table A2 Genotype count for seven SLE-associated SNPs in the MuTHER LCL data set (Grundberg et al., 2012).

SNP	Location on chromosome 6 (gene)	Minor allele	Genotype 0 count	Genotype 1 count	Genotype 2 count	Minor allele frequency
rs409558	31816126 (*MSH5*)	G	505	153	10	12.95%
rs2187668	32713862 (*HLA-DQA1*)	T	525	161	14	13.50%
rs3131379	31829012 (*MSH5*)	A	526	161	13	13.36%
rs3135391	32518965 (*HLA-DRA*)	A	503	143	10	12.42%
rs3117213	33172583 (*HLA-DPB1*)	T	376	261	57	27.02%
rs2071351	33151908 (*HLA-DPB1*)	G	477	185	28	17.46%
rs2071349	33151498 (*HLA-DPB1*)	G	535	146	12	12.27%

All genomic coordinates are from the NCBI Build 36 genomic assembly.

Table A3 Genotype count for seven SLE-associated SNPs in the HapMap CEU LCL data set.

SNP	Location on chromosome 6 (gene)	Minor allele	Genotype 0 count	Genotype 1 count	Genotype 2 count	Minor allele frequency
rs409558	31816126 (*MSH5*)	G	39	17	1	16.67%
rs2187668	32713862 (*HLA-DQA1*)	T	45	11	0	9.82%
rs3131379	31829012 (*MSH5*)	A	44	11	0	10.00%
rs3135391	32518965 (*HLA-DRA*)	A	31	21	4	25.89%
rs3117213	33172583 (*HLA-DPB1*)	T	28	20	8	32.14%
rs2071351	33151908 (*HLA-DPB1*)	G	44	11	1	11.61%
rs2071349	33151498 (*HLA-DPB1*)	G	NA	NA	NA	NA

All genomic coordinates are from the NCBI Build 36 genomic assembly. NA – no data available.

Appendix B

Table B1 Genomic location of rs409558 and proxy SNPs (r^2 > 0.80) in the Spanish (IBS) 1000 Genomes population.

SNP	r^2 with rs409558	Location on chromosome 6	Gene region
rs409558	-	31816126	intron 1 of *MSH5*
rs2293861	1.00	31819103	intron 4 of *MSH5*
rs2075788	1.00	31820160	intron 7 of *MSH5*
rs3749953	1.00	31821103	intron 9 of *MSH5*
rs3828922	1.00	31821433	intron 9 of *MSH5*
rs2075801	1.00	31836246	intron 19 of *MSH5*
rs6905572	1.00	31839860	exon 3 of *SAPCD1*
rs9469051	1.00	31849469	intron 12 of *VWA7*
rs4711277	1.00	31877663	intron 3 of *LSM2*
rs4711279	1.00	31910444	*C6orf48* promoter
rs116828153	1.00	31912625	intron 3 of *C6orf48*
rs116201952	1.00	31915468	3' UTR of *C6orf48*
rs80260075	1.00	31963679	Intergenic, between *C6orf48* and *NEU1*

All genomic coordinates are from the NCBI Build 36 genomic assembly.

Table B2 Genomic location of rs409558 and proxy SNPs (r^2 > 0.80) in the British (GBR) 1000 Genomes population.

SNP	r^2 with rs409558	Location on chromosome 6	Gene region
rs409558	-	31816126	intron 1 of *MSH5*
rs2293861	1.00	31819103	intron 4 of *MSH5*
rs2075788	1.00	31820160	intron 7 of *MSH5*
rs3828922	1.00	31821433	intron 9 of *MSH5*
rs2075801	1.00	31836246	intron 19 of *MSH5*
rs9469051	0.96	31849469	intron 12 of *VWA7*

All genomic coordinates are from the NCBI Build 36 genomic assembly.

Appendix C

Table C1 SLE-associated loci from the GWAS NHGRI database.

Reported Gene	SNP	Region	OR	P-value
AFF1	rs340630	4q21.3	1.21	8.00×10^{-9}
ARID5B, RTKN2	rs4948496	10q21.2	1.1765	5.00×10^{-11}
BANK1	rs10516487	4q24	1.38	4.00×10^{-10}
BLK	rs7812879	8p23.1	1.45	2.00×10^{-24}
C8orf13	rs13277113	8p23.1	1.39	1.00×10^{-10}
CD80	rs6804441	3q13.33	1.2658	3.00×10^{-16}
CDKN1B	rs34330	12p13.1	1.1905	5.00×10^{-12}
CREBL2	rs12822507	12p13.2	1.1628	2.00×10^{-8}
DGUOK, TET3	rs6705628	2p13.1	1.3333	7.00×10^{-17}
DRAM1	rs4622329	12q23.2	1.1905	9.00×10^{-12}
EDEM3	rs10911628	1q25.3	1.954	2.00×10^{-13}
ELF1	rs7329174	13q14.11	1.26	1.00×10^{-8}
ETS1	rs6590330	11q24.3	1.37	2.00×10^{-25}
FAM98B, TYRO3	rs11073328	15q14	1.935	1.00×10^{-14}
GPR19	rs10845606	12p13.1	1.2658	4.00×10^{-17}
IKZF1	rs4917014	7p12.2	1.39	3.00×10^{-23}
IRF5	rs4728142	7q32.1	1.43	8.00×10^{-19}
ITGAM	rs9888739	16p11.2	1.62	2.00×10^{-23}
ITGAX	rs11574637	16p11.2	1.33	3.00×10^{-11}
KIAA1542	rs4963128	11p15.5	1.28	3.00×10^{-10}
LRRC18	rs1913517	10q11.23	1.24	7.00×10^{-12}
PRDM1, ATG5	rs548234	6q21	1.25	5.00×10^{-12}
PXK	rs6445975	3p14.3	1.25	7.00×10^{-9}
RASGRP3	rs13385731	2p22.3	1.43	1.00×10^{-15}
RNF114	rs11697848	20q13.13	2.115	1.00×10^{-11}
SLC15A4	rs1385374	12q24.33	1.26	2.00×10^{-11}
SPATA8	rs8023715	15q26.2	1.812	1.00×10^{-8}
STAT4	rs7574865	2q32.2 / 2q32.3	1.51	5.00×10^{-42}
TNFAIP3	rs2230926	6q23.3	1.72	1.00×10^{-17}
TNFSF4	rs2205960	1q25.1	1.46	3.00×10^{-32}
TNIP1	rs10036748	5q33.1	1.23	2.00×10^{-9}

Reported Gene	SNP	Region	OR	P-value
TNPO3	rs10488631	7q32.1	1.829	2.00×10^{-13}
UBE2L3, HIC2	rs131654	22q11.21	1.28	1.00×10^{-16}
WDFY4	rs7097397	10q11.23	1.3	8.00×10^{-12}
ZNF184	rs10946940	6p22.1	1.45	8.00×10^{-9}

All associated loci passed a genome-wide significance threshold of 5×10^{-8}. Associations within the xMHC region, intergenic regions, and non-annotated loci are not included. OR – odds ratio.

Table C2 Rheumatoid arthritis-associated loci from the GWAS NHGRI database.

Reported Gene	SNP	Region	OR	P-value
AFF3	rs11676922	2q11.2	1.12	1.00×10^{-14}
AIRE, PFKL	rs2075876	21q22.3	1.18	4.00×10^{-9}
ANKRD55, IL6ST	rs6859219	5q11.2	1.28	1.00×10^{-11}
ANXA3	rs2867461	4q21.21	1.13	1.00×10^{-12}
ARID5B	rs10821944	10q21.2	1.16	6.00×10^{-18}
B3GNT2	rs11900673	2p15	1.11	1.00×10^{-8}
BLK	rs2736340	8p23.1	1.19	6.00×10^{-9}
C5	rs881375	9q33.2	ND	4.00×10^{-8}
C5orf30	rs26232	5q21.1	1.14	4.00×10^{-8}
CCL21	rs951005	9p13.3	1.19	4.00×10^{-10}
CCR6	rs3093024	6q27	1.19	8.00×10^{-19}
CD247	rs864537	1q24.2	ND	2.00×10^{-11}
CD40	rs4810485	20q13.12	1.18	3.00×10^{-9}
CD83	rs12529514	6p23	1.14	2.00×10^{-8}
CSF2	rs657075	5q31.1	1.12	3.00×10^{-10}
CTLA4	rs231735	2q33.2	1.17	6.00×10^{-9}
DDX6	rs10892279	11q23.3	ND	1.00×10^{-12}
ELMO1	rs11984075	7p14.1	ND	5.00×10^{-8}
GATSL3	rs1043099	22q12.2	1.19	7.00×10^{-9}
IL2RA	rs706778	10p15.1	1.14	1.00×10^{-11}
IRF5	rs10488631	7q32.1	1.19	4.00×10^{-11}
LBH	rs7579944	2p23.1	ND	1.00×10^{-8}
NFKBIE	rs2233434	6p21.1	1.19	6.00×10^{-19}
OLIG3	rs10499194	6q23.3	1.33	1.00×10^{-9}
PADI4	rs2240335	1p36.13	1.50	2.00×10^{-8}

Reported Gene	SNP	Region	OR	P-value
PDE2A, ARAP1	rs3781913	11q13.4	1.12	6.00×10^{-10}
PLD4	rs2841277	14q32.33	1.15	2.00×10^{-14}
PSMA4	rs12901682	15q25.1	135.02	4.00×10^{-8}
PTPN2	rs1893217	18p11.21	ND	5.00×10^{-12}
PTPN22	rs2476601	1p13.2	1.94	9.00×10^{-74}
PXK	rs13315591	3p14.3	1.29	5.00×10^{-8}
RBPJ	rs874040	4p15.2	1.14	1.00×10^{-16}
REL	rs13017599	2p16.1	1.21	2.00×10^{-12}
RTKN2	rs3125734	10q21.2	1.20	5.00×10^{-9}
SH2B3	rs653178	12q24.12	ND	3.00×10^{-19}
SPRED2	rs934734	2p14	1.13	5.00×10^{-10}
SPSB1	rs11121380	1p36.22	11.11	5.00×10^{-8}
STAT4	rs7574865	2q32.3	ND	4.00×10^{-10}
TNFAIP3	rs6920220	6q23.3	1.22	9.00×10^{-13}
TRAF1	rs1953126	9q33.2	ND	4.00×10^{-11}
UBASH3A	rs11203203	21q22.3	ND	1.00×10^{-8}
UBE2L3	rs2298428	22q11.21	ND	3.00×10^{-10}

All associated loci passed a genome-wide significance threshold of 5×10^{-8}. Associations within the xMHC region, intergenic regions, and non-annotated loci are not included. OR – odds ratio, ND – no data available.

Table C3 Sjögren's syndrome-associated loci from the GWAS NHGRI database.

Reported Gene	SNP	Region	OR	P-value
GTF2I, GTF2IRD1	rs117026326	7q11.23	2.20	1.00×10^{-53}
STAT4	rs10168266	2q32.3	1.44	2.00×10^{-17}
TNFAIP3	rs5029939	6q23.3	1.67	8.00×10^{-9}

All associated loci passed a genome-wide significance threshold of 5×10^{-8}. Associations within the xMHC region, intergenic regions, and non-annotated loci are not included. OR – odds ratio.

Table C4 Idiopathic inflammatory myopathies-associated loci from the GWAS NHGRI database.

Reported Gene	SNP	Region	OR	P-value
PTPN22	rs2476601	1p13.2	1.32	7.22×10^{-9}

All associated loci passed a genome-wide significance threshold of 5×10^{-8}. Associations within the xMHC region, intergenic regions, and non-annotated loci are not included. OR – odds ratio.

Table C5 Multiple sclerosis-associated loci from the GWAS NHGRI database.

Reported Gene	SNP	Region	OR	P-value
AGAP2, CYP27B1	rs12368653	12q14.1	1.10	2.00×10^{-9}
AHI1	rs11154801	6q23.3	1.13	1.00×10^{-13}
BACH2	rs12212193	6q15	1.09	4.00×10^{-8}
BATF	rs2300603	14q24.3	1.11	2.00×10^{-8}
C1orf106, KIF21B	rs7522462	1q32.1	1.11	2.00×10^{-9}
C3orf1, TMEM39A	rs2293370	3q13.33	1.13	3.00×10^{-9}
CBLB	rs9657904	3q13.11	1.40	2.00×10^{-10}
CD5	rs650258	11q12.2	1.12	2.00×10^{-11}
CD58	rs1335532	1p13.1	1.22	3.00×10^{-16}
CD6	rs4939490	11q12.2	1.30	1.00×10^{-9}
CD86	rs9282641	3q13.33	1.21	1.00×10^{-11}
CLEC16A	rs7200786	16p13.13	1.15	9.00×10^{-17}
CLECL1	rs10466829	12p13.31	1.09	1.00×10^{-8}
CYP24A1	rs2248359	20q13.2	1.12	3.00×10^{-11}
DKKL1	rs2303759	19q13.33	1.11	5.00×10^{-9}
EOMES	rs11129295	3p24.1	1.11	1.00×10^{-9}
EVI5	rs11810217	1p22.1	1.15	6.00×10^{-15}
EXTL2, VCAM1	rs12048904	1p21.2	1.09	4.00×10^{-8}
GPR65	rs2119704	14q31.3	1.22	2.00×10^{-10}
HHEX	rs7923837	10q23.33	1.10	5.00×10^{-9}
IL12B	rs2546890	5q33.3	1.11	1.00×10^{-11}
IL22RA2	rs17066096	6q23.3	1.14	6.00×10^{-13}
IL2RA	rs7090512	10p15.1	1.19	5.00×10^{-20}
IL7R	rs6897932	5p13.2	1.11	2.00×10^{-8}
IRF8	rs17445836	16q24.1	1.25	4.00×10^{-9}
KIF1B	rs10492972	1p36.22	1.34	3.00×10^{-10}
MALT1	rs7238078	18q21.32	1.12	3.00×10^{-9}
MAPK1	rs2283792	22q11.21	1.10	5.00×10^{-9}
MERTK	rs17174870	2q13	1.11	1.00×10^{-8}
METTL1	rs703842	12q14.1	1.23	5.00×10^{-11}
MLANA	rs2150702	9p24.1	1.16	3.00×10^{-8}
MMEL1	rs4648356	1p36.32	1.14	1.00×10^{-14}
MPV17L2	rs874628	19p13.11	1.11	1.00×10^{-8}
MYC	rs4410871	8q24.21	1.11	8.00×10^{-9}

Reported Gene	SNP	Region	OR	P-value
NCOA5, CD40	rs2425752	20q13.12	1.11	5.00×10^{-10}
NFKBIZ	rs771767	3q12.3	1.10	9.00×10^{-9}
ODF3B, SCO2	rs140522	22q13.33	1.10	2.00×10^{-8}
OLIG3	rs13192841	6q23.3	1.10	1.00×10^{-8}
PLEK	rs7595037	2p13.3	1.11	5.00×10^{-11}
PTGER4	rs4613763	5p13.1	1.20	3.00×10^{-16}
PTPRK, THEMIS	rs802734	6q22.33	1.10	6.00×10^{-9}
PVT1	rs2019960	8q24.21	1.12	5.00×10^{-9}
RGS1	rs1323292	1q31.2	1.12	2.00×10^{-8}
RNASEL	rs533259	1q25.3	ND	6.00×10^{-9}
SLC15A2	rs4285028	3q13.33	1.10	2.00×10^{-8}
SLC30A7	rs11581062	1p21.2	1.12	3.00×10^{-10}
SP140	rs10201872	2q37.1	1.14	2.00×10^{-10}
STAT3	rs9891119	17q21.2	1.11	2.00×10^{-10}
TAGAP	rs1738074	6q25.3	1.13	7.00×10^{-15}
THADA	rs6718520	2p21	1.17	3.00×10^{-8}
TNFRSF1A	rs1800693	12p13.31	1.12	4.00×10^{-14}
TNFSF14	rs1077667	19p13.3	1.16	9.00×10^{-14}
ZFP36L1	rs4902647	14q24.1	1.11	9.00×10^{-12}
ZMIZ1	rs1250550	10q22.3	1.10	6.00×10^{-9}
ZNF767, ZNF746	rs354033	7q36.1	1.11	5.00×10^{-9}

All associated loci passed a genome-wide significance threshold of 5×10^{-8}. Associations within the xMHC region, intergenic regions, and non-annotated loci are not included. OR – odds ratio, ND – no data available.

Table C6 Inflammatory bowel disease-associated loci from the GWAS NHGRI database.

Reported Gene	SNP	Region	OR	P-value
ADCY3	rs6545800	2p23.3	1.11	6.00×10^{-16}
ATF4, TAB1, APOBEC3G	rs2413583	22q13.1	1.21	4.00×10^{-33}
C11orf9, FADS1, FADS2	rs4246215	11q12.2	1.08	2.00×10^{-15}
C1orf53	rs2488389	1q31.3	1.12	8.00×10^{-13}
CARD15	rs2076756	16q12.1	ND	5.00×10^{-10}
CARD9, PMPCA, SDCCAG3, INPP5E	rs10781499	9q34.3	1.19	4.00×10^{-56}
CCDC88B, RPS6KA4, TRPT1, FLRT1	rs559928	11q13.1	1.10	4.00×10^{-11}
CCL13, CCL2, CCL11	rs3091316	17q12	1.12	1.00×10^{-26}
CCR6, RPS6KA2, RNASET2	rs1819333	6q27	1.08	7.00×10^{-21}
CD226	rs727088	18q22.2	1.08	5.00×10^{-9}
CD40, MMP9, PLTP	rs1569723	20q13.12	1.09	1.00×10^{-13}
CD48, SLAMF1, ITLN1, CD244, F11R, USF1, SLAMF7, ARHGAP30	rs4656958	1q23.3	1.06	7.00×10^{-9}
CD6, CD5, PTGDR2	rs11230563	11q12.2	1.09	9.00×10^{-13}
CEBPB	rs913678	20q13.13	1.06	5.00×10^{-8}
CEBPG	rs17694108	19q13.11	1.10	6.00×10^{-15}
CISD1, IPMK	rs2790216	10q21.1	1.07	8.00×10^{-9}
CNTF, LPXN	rs10896794	11q12.1	1.08	7.00×10^{-10}
CREM	rs11010067	10p11.21	1.12	2.00×10^{-25}
CRTC3	rs7495132	15q26.1	1.13	9.00×10^{-11}
CXCL5, CXCL1, CXCL3, IL8, CXCL6, PF4, CXCL2, PF4V1	rs2472649	4q13.3	1.10	3.00×10^{-8}
CXCR5	rs630923	11q23.3	1.07	7.00×10^{-9}
DAP	rs2930047	5p15.2	1.07	1.00×10^{-8}
DNMT3B	rs4911259	20q11.21	1.08	1.00×10^{-9}
DOK3	rs12654812	5q35.3	1.07	2.00×10^{-8}
EPO	rs1734907	7q22.1	1.11	2.00×10^{-13}
ERAP2, ERAP1, LNPEP	rs1363907	5q15	1.07	6.00×10^{-13}
FCGR2A, FCGR2B, FCGR3A, HSPA6, FCGR3B, FCRLA	rs1801274	1q23.3	1.12	2.00×10^{-38}
FOS, MLH3	rs4899554	14q24.3	1.08	3.00×10^{-8}
FOSL2, BRE	rs925255	2p23.2	1.09	3.00×10^{-15}
GPR183, GPR18	rs9557195	13q32.3	1.11	2.00×10^{-14}
GPR35	rs3749171	2q37.3	1.14	3.00×10^{-21}
GPR65, GALC	rs8005161	14q31.3	1.15	2.00×10^{-14}

Reported Gene	SNP	Region	OR	P-value
HCK	rs6142618	20q11.21	1.07	6.00×10^{-10}
ICOSLG	rs7282490	21q22.3	1.11	2.00×10^{-26}
IFIH1	rs2111485	2q24.2	1.07	2.00×10^{-8}
IFNG, IL26, IL22	rs7134599	12q15	1.10	9.00×10^{-32}
IKZF3, ZPBP2, GSDMB, ORMDL3, GSDMA	rs12946510	17q12	1.16	4.00×10^{-38}
IL10, IL20, IL19, IL24, PIGR, MAPKAPK2, FAIM3, RASSF5	rs3024505	1q32.1	1.21	7.00×10^{-42}
IL12B	rs6871626	5q33.3	1.18	1.00×10^{-42}
IL1R2, IL18RAP, IL18R1, IL1R1, IL1RL1, IL1RL2	rs917997	2q12.1	1.10	3.00×10^{-20}
IL2, IL21	rs7657746	4q27	1.12	3.00×10^{-13}
IL23R, IL12RB2	rs11209026	1p31.3	2.01	8.00×10^{-161}
IL2RA, IL15RA	rs12722515	10p15.1	1.10	4.00×10^{-10}
IRF1, IL13, CSF2, SLC22A4, IL4, IL3, IL5, PDLIM4, SLC22A5, ACSL6	rs2188962	5q31.1	1.16	1.00×10^{-52}
IRF8	rs10521318	16q24.1	1.16	1.00×10^{-9}
JAK2	rs10758669	9p24.1	1.17	8.00×10^{-45}
KIF21B	rs7554511	1q32.1	1.16	1.00×10^{-32}
LIF, OSM, MTMR3	rs2412970	22q12.2	1.08	3.00×10^{-14}
LOH12CR1	rs11612508	12p13.2	1.06	1.00×10^{-8}
LRRK2, MUC19	rs11564258	12q12	1.33	6.00×10^{-29}
MAP3K8	rs1042058	10p11.23	1.08	6.00×10^{-11}
MAPK1, YDJC, UBE2L3, RIMBP3, CCDC116	rs2266959	22q11.21	1.11	1.00×10^{-16}
MST1, PFKFB4, MST1R, UCN2, GPX1, IP6K2, BSN, IP6K1, USP4	rs3197999	3p21.31	1.18	1.00×10^{-47}
NFIL3	rs4743820	9q22.31	1.06	4.00×10^{-9}
NKX2-3	rs4409764	10q24.2	1.18	1.00×10^{-54}
NLRP7, NLRP2, KIR2DL1, LILRB4	rs11672983	19q13.42	1.09	7.00×10^{-11}
NOD2	rs5743289	16q12.1	1.46	4.00×10^{-10}
PHACTR2	rs12199775	6q24.2	1.13	2.00×10^{-8}
PRKCB	rs7404095	16p12.2	1.06	1.00×10^{-9}
PSMG1	rs2836878	21q22.2	1.41	4.00×10^{-12}
PTGER4	rs11742570	5p13.1	1.20	2.00×10^{-82}
RABEP2, IL27, EIF3C, SULT1A1, SULT1A2, NUPR1	rs26528	16p11.2	1.10	1.00×10^{-21}
REL, C2orf74, KIAA1841, AHSA2	rs7608910	2p16.1	1.14	9.00×10^{-32}
RELA, FOSL1, CTSW, SNX32	rs2231884	11q13.1	1.08	3.00×10^{-10}

Reported Gene	SNP	Region	OR	P-value
RORC	rs4845604	1q21.3	1.14	4.00×10^{-16}
SLC11A1, CXCR2, CXCR1, PNKD, ARPC2, TMBIM1, CTDSP1	rs2382817	2q35	1.07	4.00×10^{-12}
SMAD3	rs17293632	15q22.33	1.07	6.00×10^{-16}
SMAD7	rs7240004	18q21.1	1.06	1.00×10^{-9}
SMURF1	rs9297145	7q22.1	1.08	8.00×10^{-12}
SOCS1, LITAF, RMI2	rs529866	16p13.13	1.12	2.00×10^{-16}
SPRED2	rs6740462	2p14	1.08	2.00×10^{-8}
SPRY4, NDFIP1	rs6863411	5q31.3	1.09	4.00×10^{-14}
STAT1, STAT4	rs1517352	2q32.3	1.08	3.00×10^{-11}
STAT3, STAT5B, STAT5A	rs12942547	17q21.2	1.10	6.00×10^{-22}
TNFAIP3	rs6920220	6q23.3	1.10	1.00×10^{-21}
TNFRSF18, TNFRSF4	rs12103	1p36.33	1.10	8.00×10^{-13}
TNFRSF6B	rs2315008	20q13.33	1.36	9.00×10^{-15}
TNFRSF6B, LIME1, SLC2A4RG, ZGPAT	rs6062504	20q13.33	1.10	1.00×10^{-23}
TNFRSF9	rs35675666	1p36.23	1.11	1.00×10^{-15}
TNFSF15	rs2006996	9q32	1.75	4.00×10^{-16}
TNFSF8, TNC	rs4246905	9q32	1.14	3.00×10^{-32}
TNIP1, IRGM, ZNF300P1	rs11741861	5q33.1	1.25	3.00×10^{-37}
TNNI2, LSP1	rs907611	11p15.5	1.07	3.00×10^{-10}
TRAF3IP2, FYN, REV3L	rs3851228	6q21	1.15	1.00×10^{-13}
TRIB1	rs921720	8q24.13	1.08	8.00×10^{-20}
TSPAN14, C10orf58	rs6586030	10q23.1	1.12	9.00×10^{-16}
TUBD1, RPS6KB1	rs1292053	17q23.1	1.08	9.00×10^{-13}
TYK2, PPAN-P2RY11, ICAM1	rs11879191	19p13.2	1.14	2.00×10^{-18}
UBQLN4, RIT1, MSTO1	rs670523	1q22	1.06	6.00×10^{-11}
VDR	rs11168249	12q13.11	1.05	8.00×10^{-9}
ZFP36L1	rs194749	14q24.1	1.08	3.00×10^{-10}
ZNF831, CTSZ	rs259964	20q13.32	1.09	1.00×10^{-12}
ZPBP, IKZF1	rs1456896	7p12.2	1.09	7.00×10^{-15}

All associated loci passed a genome-wide significance threshold of 5×10^{-8}. Associations within the xMHC region, intergenic regions, and non-annotated loci are not included. OR – odds ratio, ND – no data available.

Table C7 Type 1 diabetes-associated loci from the GWAS NHGRI database.

Reported Gene	SNP	Region	OR	P-value
BACH2	rs11755527	6q15	1.13	5.00×10^{-12}
C10orf59	rs10509540	10q23.31	1.33	1.00×10^{-28}
C1QTNF6	rs229541	22q12.3	1.11	2.00×10^{-8}
C6orf173	rs9388489	6q22.32	1.17	4.00×10^{-13}
CD226	rs763361	18q22.2	1.16	1.00×10^{-8}
CD69	rs4763879	12p13.31	1.09	2.00×10^{-11}
COBL	rs4948088	7p12.1	1.30	4.00×10^{-8}
CTLA4	rs3087243	2q33.2	ND	1.00×10^{-15}
CTSH	rs3825932	15q25.1	1.16	3.00×10^{-15}
CUX2	rs1265564	12q24.12	1.45	1.00×10^{-16}
DLK1, MEG3, RTL1, DIO3	rs941576	14q32.2	1.11	1.00×10^{-10}
EFR3B, 3NCOA1, C2orf79, CENPO, ADCY3, DNAJC27, POMC, DNMT3A	rs478222	2p23.3	1.22	4.00×10^{-9}
ERBB3	rs2292239	12q13.2	ND	2.00×10^{-25}
GLIS3	rs7020673	9p24.2	1.14	5.00×10^{-12}
IFIH1	rs1990760	2q24.2	1.18	2.00×10^{-11}
IL10	rs3024505	1q32.1	1.19	2.00×10^{-9}
IL2	rs4505848	4q27	ND	5.00×10^{-13}
IL27	rs4788084	16p11.2	1.09	3.00×10^{-13}
IL2RA	rs12251307	10p15.1	ND	1.00×10^{-13}
INS	rs7111341	11p15.5	ND	4.00×10^{-48}
KIAA0350, CLEC16A	rs12708716	16p13.13	1.23	3.00×10^{-18}
LMO7	rs539514	13q22.2	1.43	6.00×10^{-11}
ORMDL3	rs2290400	17q12	1.15	6.00×10^{-13}
PHTF1	rs6679677	1p13.2	1.89	8.00×10^{-24}
PRKCQ	rs947474	10p15.1	1.10	4.00×10^{-9}
PTPN2	rs1893217	18p11.21	ND	4.00×10^{-15}
PTPN22	rs2476601	1p13.2	ND	9.00×10^{-85}
RAB5B, SUOX, IKZF4, CDK2	rs1701704	12q13.2	1.25	9.00×10^{-10}
SH2B3	rs3184504	12q24.12	ND	3.00×10^{-27}
LNK, TRAFD1, PTPN11, C12orf30	rs17696736	12q24.13	1.34	2.00×10^{-14}
TYK2	rs2304256	19p13.2	1.16	4.00×10^{-9}
UBASH3A	rs11203203	21q22.3	ND	2.00×10^{-9}
WDR27, C6orf120, PHF10, TCTE3, C6orf208, LOC154449, DLL1, FAM120B, PSMB1, TBP, PCD2	rs924043	6q27	1.35	8.00×10^{-9}

All associated loci passed a genome-wide significance threshold of 5×10^{-8}. Associations within the xMHC region, intergenic regions, and non-annotated loci are not included. OR – odds ratio, ND – no data available.

Table C8 Psoriasis-associated loci from the GWAS NHGRI database.

Reported Gene	SNP	Region	OR	P-value
ERAP1	rs27524	5q15	1.13	3.00×10^{-11}
FBXL19, POL3S	rs10782001	16p11.2	1.16	9.00×10^{-10}
IFIH1	rs17716942	2q24.2	1.29	1.00×10^{-13}
IL12B	rs2082412	5q33.3	1.44	2.00×10^{-28}
IL13	rs20541	5q31.1	1.27	5.00×10^{-15}
IL23A, STAT2	rs2066808	12q13.3	1.34	1.00×10^{-9}
IL23R	rs2201841	1p31.3	1.13	3.00×10^{-8}
LCE3A	rs4085613	1q21.3	1.32	7.00×10^{-30}
LCE3D	rs4112788	1q21.3	1.29	3.00×10^{-10}
NFKBIA	rs8016947	14q13.2	1.19	2.00×10^{-11}
NOS2	rs4795067	17q11.2	1.19	4.00×10^{-11}
PSMA6	rs12586317	14q13.2	1.15	2.00×10^{-8}
REL	rs702873	2p16.1	1.12	4.00×10^{-9}
SPATA2	rs495337	20q13.13	1.25	1.00×10^{-8}
TNFAIP3	rs610604	6q23.3	1.19	9.00×10^{-12}
TNIP1	rs17728338	5q33.1	1.59	1.00×10^{-20}
TRAF3IP2	rs240993	6q21	1.25	5.00×10^{-20}
TYK2	rs12720356	19p13.2	1.4	4.00×10^{-11}

All associated loci passed a genome-wide significance threshold of 5×10^{-8}. Associations within the xMHC region, intergenic regions, and non-annotated loci are not included. OR – odds ratio.

Table C9 Osteoarthritis-associated loci from the GWAS NHGRI database.

Reported Gene	SNP	Region	OR	P-value
DOT1L	rs12982744	19p13.3	0.09	1.00×10^{-11}
DUS4L	rs4730250	7q22.3	1.15	6.00×10^{-8}
GNL3, GLT8D1	rs11177	3p21.1	1.09	5.00×10^{-9}
MCF2L	rs11842874	13q34	1.17	2.00×10^{-8}

All associated loci passed a genome-wide significance threshold of 5×10^{-8}. Associations within the xMHC region, intergenic regions, and non-annotated loci are not included. OR – odds ratio.

Table C10 Type 2 diabetes-associated loci from the GWAS NHGRI database.

Reported Gene	SNP	Region	OR	P-value
ADAMTS9	rs4607103	3p14.1	1.09	1.00×10^{-8}
ANK1	rs515071	8p11.21	1.18	1.00×10^{-8}
AP3S2	rs2028299	15q26.1	1.1	2.00×10^{-11}
ARF5, SND1	rs10229583	7q32.1	1.14	2.00×10^{-10}
BCL11A	rs243021	2p16.1	1.08	3.00×10^{-15}
C2CD4A, C2CD4B	rs7172432	15q22.2	1.11	9.00×10^{-14}
C6orf57	rs1048886	6q13	1.54	3.00×10^{-8}
CAMK1D	rs12779790	10p13	1.11	1.00×10^{-10}
CDC123	rs11257655	10p13	1.15	7.00×10^{-9}
CDKAL1	rs10440833	6p22.3	1.25	2.00×10^{-22}
CDKN2A, CDKN2B	rs2383208	9p21.3	1.34	2.00×10^{-29}
CENTD2	rs1552224	11q13.4	1.14	1.00×10^{-22}
CHCHD9	rs13292136	9q21.31	1.11	3.00×10^{-8}
DUSP9	rs5945326	Xq28	1.18	7.00×10^{-16}
FAM58A	rs12010175	Xq28	1.21	2.00×10^{-9}
FITM2, R3HDML, HNF4A	rs6017317	20q13.12	1.09	1.00×10^{-11}
FTO	rs9939609	16q12.2	1.25	1.00×10^{-20}
GCC1, PAX4	rs6467136	7q32.1	1.11	5.00×10^{-11}
GLIS3	rs7041847	9p24.2	1.1	2.00×10^{-14}
GPSM1	rs11787792	9q34.3	1.15	2.00×10^{-10}
GRB14	rs3923113	2q24.3	1.09	1.00×10^{-8}
GRK5	rs10886471	10q26.11	1.12	7.00×10^{-9}
HHEX	rs1111875	10q23.33	1.21	7.00×10^{-12}
HMG20A	rs7178572	15q24.3	1.09	7.00×10^{-11}
HMGA2	rs1531343	12q14.3	1.1	4.00×10^{-9}
HNF1A	rs7305618	12q24.31	1.14	2.00×10^{-8}
HNF1B	rs4430796	17q12	1.19	2.00×10^{-11}
HNF4A	rs4812829	20q13.12	1.09	3.00×10^{-10}
IDE	rs5015480	10q23.33	1.18	1.00×10^{-15}
IGF2BP2	rs1470579	3q27.2	1.08	2.00×10^{-19}
IRS1	rs7578326	2q36.3	1.11	5.00×10^{-20}
JAZF1	rs864745	7p15.1	1.1	5.00×10^{-14}
KCNJ11	rs5215	11p15.1	1.14	5.00×10^{-11}
KCNK16	rs1535500	6p21.2	1.08	2.00×10^{-8}

Reported Gene	SNP	Region	OR	P-value
KCNQ1	rs2237892	11p15.4	1.4	2.00×10^{-42}
KLF14	rs972283	7q32.3	1.07	2.00×10^{-10}
LAMA1	rs8090011	18p11.31	1.13	8.00×10^{-9}
MAEA	rs6815464	4p16.3	1.13	2.00×10^{-20}
MGC21675	rs7656416	4p16.3	1.15	1.00×10^{-8}
MIR129, LEP	rs791595	7q32.1	1.17	3.00×10^{-13}
MTNR1B	rs1387153	11q14.3	1.09	8.00×10^{-15}
NOTCH2, ADAM30	rs10923931	1p12	1.13	4.00×10^{-8}
PEPD	rs3786897	19q13.11	1.1	1.00×10^{-8}
PRC1	rs8042680	15q26.1	1.07	2.00×10^{-10}
PSMD6	rs831571	3p14.1	1.09	8.00×10^{-11}
PTPRD	rs17584499	9p24.1	1.57	9.00×10^{-10}
RASGRP1	rs7403531	15q14	1.1	4.00×10^{-9}
RBM43, RND3	rs7560163	2q23.3	1.33	7.00×10^{-9}
RBMS1, ITGB6	rs7593730	2q24.2	1.11	4.00×10^{-8}
SGCG, SACS	rs9552911	13q12.12	1.49	2.00×10^{-8}
SLC16A13	rs312457	17p13.1	1.2	8.00×10^{-13}
SLC30A8	rs13266634	8q24.11	1.22	2.00×10^{-14}
SPRY2	rs1359790	13q31.1	1.15	6.00×10^{-9}
SRR	rs391300	17p13.3	1.28	3.00×10^{-9}
ST6GAL1	rs16861329	3q27.3	1.09	3.00×10^{-8}
TCF7L2	rs7903146	10q25.2	1.19	9.00×10^{-75}
THADA	rs7578597	2p21	1.15	1.00×10^{-9}
TP53INP1	rs896854	8q22.1	1.06	1.00×10^{-9}
TSPAN8, LGR5	rs7961581	12q21.1	1.09	1.00×10^{-9}
VPS26A	rs1802295	10q22.1	1.08	4.00×10^{-8}
WFS1	rs1801214	4p16.1	1.13	3.00×10^{-8}
ZBED3	rs4457053	5q13.3	1.08	3.00×10^{-12}
ZFAND3	rs9470794	6p21.2	1.12	2.00×10^{-10}
ZFAND6	rs11634397	15q25.1	1.06	2.00×10^{-9}

All associated loci passed a genome-wide significance threshold of 5×10^{-8}. Associations within the xMHC region, intergenic regions, and non-annotated loci are not included. OR – odds ratio.

Table C11 Schizophrenia-associated loci from the GWAS NHGRI database.

Reported Gene	SNP	Region	OR	P-value
AKT3, CEP170	rs14403	1q43	1.10	2.00×10^{-8}
BRP44, DCAF6	rs10489202	1q24.2	1.23	1.00×10^{-8}
C10orf32, AS3MT, CALHM1, CALHM2, CALHM3, CNNM2, CYP17A1, INA, MIR1307, NT5C2, PCGF6, PDCD11, SFXN2, ST13P13, TAF5, USMG5, WBP1L	rs7085104	10q24.32	1.11	4.00×10^{-13}
C12orf65, ABCB9, ARL6IP4, CDK2AP1, MIR4304, MPHOSPH9, OGFOD2, PITPNM2, RILPL2, SBNO1, SETD8	rs11532322	12q24.31	1.09	2.00×10^{-8}
C2orf82, GIGYF2, KCNJ13, NGEF	rs778371	2q37.1	1.09	2.00×10^{-8}
CACNA1C	rs1006737	12p13.33	1.10	5.00×10^{-12}
CACNB2, NSUN6	rs17691888	10p12.31	1.16	1.00×10^{-10}
CCDC68	rs12966547	18q21.2	1.40	3.00×10^{-8}
CNNM2	rs7914558	10q24.32	1.22	2.00×10^{-8}
CSMD1	rs10503253	8p23.2	1.16	2.00×10^{-8}
FONG, C2orf47, C2orf69, SPATS2L, TYW5	rs2949006	2q33.1	1.10	1.00×10^{-8}
GRIA1	rs17504622	5q33.1	1.24	3.00×10^{-9}
HHAT	rs7527939	1q32.2	2.63	6.00×10^{-9}
ITIH3, ALAS1, ALDOAP1, BAP1, C3orf78, DNAH1, GLT8D1, GLYCTK, GNL3, ITIH1, ITIH4, MIR135A1, MIRLET7G, MUSTN1, NEK4, NISCH, NT5DC2, PBRM1, PHF7, PPM1M, RFT1, SEMA3G, SFMBT1, SPCS1, STAB1, TLR9, TMEM110, TNNC1, TWF2, WDR82	rs4687552	3p21.1	1.09	1.00×10^{-8}
LSM1, WHSC1L1	rs16887244	8p11.23	1.19	1.00×10^{-10}
MAD1L1, FTSJ2, NUDT1, SNX8	rs6461049	7p22.3	1.11	6.00×10^{-13}
MAD1L1, SNX8, NUDT1, FTSJ2	rs12666575	7p22.3	1.12	2.00×10^{-9}
MAU2, CILP2, GATAD2A, GMIP, HAPLN4, LPAR2, MIR640, NCAN, NDUFA13, PBX4, SUGP1, TM6SF2, TSSK6, YJEFN3	rs2905424	19p13.11	1.09	3.00×10^{-9}
MIR137	rs1625579	1p21.3	1.12	2.00×10^{-11}
MIR137, DPYD	rs1198588	1p21.3	1.12	2.00×10^{-12}
MMP16	rs11995572	8q21.3	1.12	3.00×10^{-8}
NRGN	rs12807809	11q24.2	1.15	2.00×10^{-9}
NT5C2	rs11191580	10q24.33	1.20	3.00×10^{-8}
NT5C2, CNNM2	rs11191580	10q24.33	1.23	2.00×10^{-9}
PCGEM1	rs17662626	2q32.3	1.20	5.00×10^{-8}
QPCT, C2orf56, CEBPZ, PRKD3, SULT6B1	rs2373000	2p22.2	1.09	7.00×10^{-9}

Reported Gene	SNP	Region	OR	P-value
RENBP, ARHGAP4, MECP2	rs2269372	Xq28	1.31	4.00×10^{-8}
SDCCAG8	rs1538774	1q43	1.09	3.00×10^{-8}
SLC17A1, SLC17A3, BTN3A2, BTN2A2, BTN3A1, HIST1H2AG, HIST1H2BJ, PRSS16, POM121L2, ZNF184	rs13194053	6p22.1	1.28	1.00×10^{-8}
SLCO6A1	rs6878284	5q21.1	1.09	9.00×10^{-9}
SNX19	rs7940866	11q25	1.09	2.00×10^{-9}
TCF4	rs1261117	18q21.2	1.60	3.00×10^{-10}
TCF4	rs9960767	18q21.2	1.23	4.00×10^{-9}
TCF4	rs4801131	18q21.2	1.08	1.00×10^{-8}
TSNARE1	rs4129585	8q24.3	1.09	2.00×10^{-10}
TSPAN18	rs11038167	11p11.2	1.29	1.00×10^{-11}
ZEB2	rs12991836	2q22.3	1.08	1.00×10^{-8}
ZSWIM6, C5orf43	rs171748	5q12.1	1.08	4.00×10^{-8}

All associated loci passed a genome-wide significance threshold of 5×10^{-8}. Associations within the xMHC region, intergenic regions, and non-annotated loci are not included. OR – odds ratio.

Table C12 Bipolar disorder-associated loci from the GWAS NHGRI database.

Reported Gene	SNP	Region	OR	P-value
ANK3	rs4948418	10q21.2	ND	4.00×10^{-10}
CANCNA1C	rs4765913	12p13.33	1.14	2.00×10^{-8}
DGKH	rs1012053	13q14.11	1.59	2.00×10^{-8}
LMAN2L	rs2271893	2q11.2	ND	2.00×10^{-10}
NCAN	rs1064395	19p13.11	1.17	2.00×10^{-9}
ODZ4	rs12576775	11q14.1	1.14	4.00×10^{-8}
PTGFR	rs4650608	1p31.1	ND	8.00×10^{-9}
RP11-252P19.1	rs1039002	6q27	ND	2.00×10^{-8}
TRANK1	rs9834970	3p22.2	ND	1.00×10^{-12}

All associated loci passed a genome-wide significance threshold of 5×10^{-8}. Associations within the xMHC region, intergenic regions, and non-annotated loci are not included. OR – odds ratio, ND – no data available.

Table C13 Hypertension-associated loci from the GWAS NHGRI database.

Reported Gene	SNP	Region	OR	P-value
ATP2B1	rs2681472	12q21.33	0.15	2.00×10^{-11}
EDN3, GNAS	rs6015450	20q13.32	0.11	4.00×10^{-14}
FLJ32810, TMEM133	rs633185	11q22.1	0.07	5.00×10^{-11}
GPR39	rs13420028	2q21.2	ND	1.00×10^{-10}
MACROD2	rs200752	20p12.1	ND	7.00×10^{-9}
MYO6	rs3798440	6q14.1	ND	3.00×10^{-10}
NOV	rs2469997	8q24.12	ND	3.00×10^{-16}
NPR3, C5orf23	rs1173771	5p13.3	0.06	3.00×10^{-10}
PLCE1	rs932764	10q23.33	0.06	9.00×10^{-9}
RANBP3L	rs7735940	5p13.2	ND	5.00×10^{-13}
UMOD	rs13333226	16p12.3	1.15	4.00×10^{-11}
ZFAT	rs7827545	8q24.22	ND	2.00×10^{-44}

All associated loci passed a genome-wide significance threshold of 5×10^{-8}. Associations within the xMHC region, intergenic regions, and non-annotated loci are not included. OR – odds ratio, ND – no data available.

Appendix D

Table D1 Genes associated with idiopathic SLE.

Reported Gene	SNP	Region	OR	P-value	Reference
STAT4	rs11889341	2q32	1.73	5.59×10^{-122}	Bentham et al. (2015)
NCF2, SMG7	rs17849501	1q25.3	2.1	3.45×10^{-88}	Bentham et al. (2015)
TNIP1	rs10036748	5q32	1.38	1.27×10^{-45}	Bentham et al. (2015)
TNFAIP3	rs6932056	6q23.3	1.83	1.97×10^{-31}	Bentham et al. (2015)
PTPN22	rs2476601	1p13	1.43	1.10×10^{-28}	Bentham et al. (2015)
MIR146A	rs2431697	5q33.3	1.26	8.01×10^{-28}	Bentham et al. (2015)
CD44	rs2732549	11p13	1.24	1.20×10^{-23}	Bentham et al. (2015)
DHCR7, NADSYN1	rs3794060	11q13.4	1.23	1.32×10^{-20}	Bentham et al. (2015)
TNFSF4 (OX40L)	rs704840	1q25.1	1.22	3.12×10^{-19}	Bentham et al. (2015)
IRF5, TNPO3	rs12537284	7q32	1.54	3.61×10^{-19}	International Consortium for Systemic Lupus Erythematosus et al. (2008)
IRF8	rs11644034	16q24.1	1.25	9.58×10^{-18}	Bentham et al. (2015)
GPR19	rs10845606	12p13	0.79	3.80×10^{-17}	Yang et al. (2013)
BANK1	rs10028805	4q24	1.2	4.31×10^{-17}	Bentham et al. (2015)
TET3, DGUOK	rs6705628	2p13	0.75	6.90×10^{-17}	Yang et al. (2013)
CIITA, SOCS1	rs9652601	16p13.13	1.21	7.42×10^{-17}	Bentham et al. (2015)
UBE2L3, HIC2	rs463426	22q11.21	0.78	1.48×10^{-16}	Han et al. (2009)
CD80	rs6804441	3q13.33	0.79	2.50×10^{-16}	Boackle (2013)
SKP1, TCF7	rs7726414	5q31.1	1.45	4.44×10^{-16}	Bentham et al. (2015)
RASGRP3	rs13385731	2p24.1	0.7	1.30×10^{-15}	Vaughn et al. (2012)
IRAK1, MECP2	rs1734787	Xq28	1.31	1.78×10^{-15}	Bentham et al. (2015)
CSK	rs2289583	15q24.1	1.19	6.22×10^{-15}	Bentham et al. (2015)
ABHD6, PXK	rs9311676	3p14.3	1.17	3.06×10^{-14}	Bentham et al. (2015)
ATG5, PRDM1	rs6568431	6q21	1.21	5.04×10^{-14}	Bentham et al. (2015)
IKZF1	rs4917014	7p12.2	1.18	6.39×10^{-14}	Bentham et al. (2015)
IKZF2	rs3768792	2q34	1.24	1.21×10^{-13}	Bentham et al. (2015)
SLC15A4	rs1059312	12q24.32	1.17	1.48×10^{-13}	Bentham et al. (2015)
TYK2	rs2304256	19p13.2	1.24	3.50×10^{-13}	Bentham et al. (2015)
FCGR2A, FCGR2B, FCGR3B, FCGR3A	rs1801274	1q23.3	1.16	1.04×10^{-12}	Bentham et al. (2015)
CDKN1B	rs34330	12p13	0.84	4.80×10^{-12}	Yang et al. (2013)
WDFY4, LRRC18	rs1913517	10q11.23	1.24	7.20×10^{-12}	Vaughn et al. (2012)

Reported Gene	SNP	Region	OR	P-value	Reference
DRAM1	rs4622329	12q23	0.84	9.40×10^{-12}	Boackle (2013)
IFIH1	rs2111485	2q24	1.15	1.27×10^{-11}	Bentham et al. (2015)
KIAA0319L	rs2275247	1p34.3	1.49	3.31×10^{-11}	Martin et al. (2013)
ARID5B, RTKN2	rs4948496	10q21	0.85	5.10×10^{-11}	Yang et al. (2013)
TREX1	rs3135945	3p21.3	44.65	8.50×10^{-11}	Vaughn et al. (2012)
JAZF1	rs849142	7p15.1	1.14	8.61×10^{-11}	Bentham et al. (2015)
BLK, C8orf13 (FAM167A)	rs13277113	8p23.1	1.39	1.00×10^{-10}	Hom et al. (2008)
ETS1, FLI1	rs7941765	11q23.3	1.14	1.35×10^{-10}	Bentham et al. (2015)
KIAA1542 (PHRF1)	rs4963128	11p15.5	0.78	3.00×10^{-10}	International Consortium for Systemic Lupus Erythematosus et al. (2008)
IKZF3, ZPBP2	rs1453560	17q21	1.23-1.37	3.48×10^{-10}	Boackle (2013)
ICAM1, ICAM4, ICAM5	rs3093030	19p13	1.16	4.88×10^{-10}	Boackle (2013)
CXorf21	rs887369	Xp21.2	1.15	5.26×10^{-10}	Bentham et al. (2015)
RAD51B	rs4902562	14q24.1	1.14	6.15×10^{-10}	Bentham et al. (2015)
TLR7	rs3853839	Xp22.3	1.67	6.50×10^{-10}	Vaughn et al. (2012)
IRF7	rs12802200	11p15.5	1.23	8.81×10^{-10}	Bentham et al. (2015)
LYST	rs9782955	1q42.3 1	1.16	1.25×10^{-9}	Bentham et al. (2015)
PRKCB	rs16972959	16p11.2	0.81	1.40×10^{-9}	Vaughn et al. (2012)
IL12A	rs564799	3q25.33	1.14	1.54×10^{-9}	Bentham et al. (2015)
PLD2	rs2286672	17p13.2	1.25	2.93×10^{-9}	Bentham et al. (2015)
SH2B3	rs10774625	12q24.12	1.13	4.09×10^{-9}	Bentham et al. (2015)
IL10	rs3024505	1q24	1.17	4.64×10^{-9}	Bentham et al. (2015)
LYN	rs7829816	8q13	0.77	5.40×10^{-9}	Vaughn et al. (2012)
UHRF1BP1	rs9462027	6p21	1.14	7.55×10^{-9}	Bentham et al. (2015)
AFF1	rs340630	4q21	1.21	8.30×10^{-9}	Boackle (2013)
TMEM39A	rs1132200	3q13.33	0.73–0.83	8.62×10^{-9}	Boackle (2013)
RPEL1	rs4917385	10q24.33	0.72	1.39×10^{-8}	Alarcon-Riquelme et al. (2016)
ELF1	rs7329174	13q13	1.26	1.50×10^{-8}	Vaughn et al. (2012)
IL21	rs907715	4q26	1.16	2.20×10^{-8}	Vaughn et al. (2012)
CREBL2	rs12822507	12p13	0.86	2.22×10^{-8}	Yang et al. (2013)
SPRED2	rs6740462	2p14	1.2	2.31×10^{-8}	Bentham et al. (2015)
ITGAM, ITGAX	rs11574637	16p11.2	1.33	3.00×10^{-11}	Hom et al. (2008)

All associated loci passed a genome-wide significance threshold of 5×10^{-8}. Associations within the xMHC region are not included. OR – odds ratio.

Table D2 Genes associated with monogenic forms of SLE.

Reported Gene	Variant(s)	Region	Reference
ACP5 (TRAP)	I89T, G215R, D241N, M264K	19p13	Briggs et al. (2011)
C1QA	Homozygous deficiency	1p36.12	Bryan and Wu (2014)
C1QB	Homozygous deficiency	1p36.12	Bryan and Wu (2014)
C1QC	Homozygous deficiency	1p36.12	Bryan and Wu (2014)
C1R	Homozygous deficiency	12p13.31	Bryan and Wu (2014)
C1S	Homozygous deficiency	12p13.31	Bryan and Wu (2014)
C2	Homozygous deficiency	6p21.33	Bryan and Wu (2014)
DNASE1	K5X	16p13	Yasutomo et al. (2001)
DNASE1L3	W215GfsX2	3p14	Al-Mayouf et al. (2011)
FASL	84-bp deletion	1q24.3	Wu et al. (1996)
PRKCD	G510S	3p21.1	Belot et al. (2013)
RNASEH2A	L202S	19p13.13	Gunther et al. (2015)
RNASEH2B	K233Q	13q14.3	Gunther et al. (2015)
RNASEH2C	348 + 1 G>A	11q13.1	Gunther et al. (2015)
SAMHD1	I201N	20q11	Ravenscroft et al. (2011)
TREX1	E266G	3p21	Lee-Kirsch et al. (2007)

Appendix E

Table E1 The effect of IFN-α stimulation on the expression of type 1 interferonopathy susceptibility genes in B cells.

Gene	Transcript cluster ID	Fold change	P_{FDR}
IFIH1	**2584207**	**9.82**	**6.09×10^{-16}**
DDX58 (RIG1)	**3203086**	**8.48**	**6.00×10^{-15}**
ADAR	**2436754**	**4.57**	**6.14×10^{-15}**
ISG15	**4053534**	**45.01**	**7.39×10^{-15}**
PSMB8	**2950199**	**2.49**	**1.25×10^{-10}**
SAMHD1	**3904691**	**3.76**	**2.14×10^{-8}**
TREX1	**2621705**	**1.48**	**1.23×10^{-7}**
ACP5 (TRAP)	**3851072**	**2.01**	**2.56×10^{-2}**
RNASEH2C	3377826	1.14	6.80×10^{-2}
RNASEH2B	3489957	1.10	9.63×10^{-2}
TMEM173 (STING)	2877990	-1.07	1.14×10^{-1}
RNASEH2A	3821908	1.03	2.62×10^{-1}

Type 1 interferonopathy-associated genes were obtained from Crow and Manel (2015). Genes differentially expressed at an FDR < 5% are indicated in bold type.

Appendix F

Table F1 The effect of IFN-α stimulation on the expression of xMHC genes in B cells.

Gene	Region	Transcript cluster ID	Fold change	P_{FDR}
TAP1	class II	2950214	3.75	8.22×10^{-15}
TNF	class III	2902416	10.00	1.52×10^{-14}
HLA-DMA	class II	2950277	-2.13	1.39×10^{-12}
PPP1R10	class I	2948425	2.41	6.28×10^{-11}
PSMB8	class II	2950199	2.49	1.25×10^{-10}
HLA-DMB	class II	2950263	-2.25	7.79×10^{-10}
LOC100507463	class II	2903285	1.91	4.37×10^{-9}
HLA-F	class I	2900974	2.16	4.75×10^{-9}
TAP2	class II	2950167	2.93	6.64×10^{-9}
LTA	class III	2902407	6.46	7.09×10^{-9}
HLA-DOA	class II	2950307	-2.55	7.47×10^{-9}
NCR3	class III	2949132	-2.78	1.99×10^{-8}
ABCF1	class I	2901687	1.56	3.48×10^{-8}
GABBR1	xclass I	2947889	-2.34	3.49×10^{-8}
HLA-DOB	class II	2950145	-2.10	5.24×10^{-8}
NRM	class I	2948547	-2.57	1.07×10^{-7}
BTN2A1	xclass I	2899437	2.03	1.12×10^{-7}
TRIM26	class I	2948259	1.87	1.28×10^{-7}
ZNRD1	class I	2901333	2.54	1.62×10^{-7}
LTB	class III	2949118	-2.11	1.65×10^{-7}
FKBPL	class III	2949760	1.72	2.17×10^{-7}
BAG6	class III	2949148	1.45	3.25×10^{-7}
AGPAT1	class III	2949801	-1.49	4.52×10^{-7}
HIST1H2AK	xclass I	2947063	-2.55	5.84×10^{-7}
HIST1H1C	xclass I	2946232	-3.63	1.40×10^{-6}
HIST1H1D	xclass I	2946353	-2.76	1.87×10^{-6}
TCF19	class I	2902178	-1.61	2.57×10^{-6}
TAPBP	xclass II	2950629	1.35	3.01×10^{-6}
MSH5-SAPCD1	class III	2902633	-1.59	4.64×10^{-6}
HSPA1B	class III	2902707	3.00	4.97×10^{-6}
MICB	class I	2902348	2.06	5.10×10^{-6}

Gene	Region	Transcript cluster ID	Fold change	P_{FDR}
DHX16	class I	2948485	1.38	5.29×10^{-6}
CLIC1	class III	2949330	2.08	6.43×10^{-6}
ZNF184	xclass I	2946845	1.59	9.71×10^{-6}
RGL2	xclass II	2950590	-1.68	1.13×10^{-5}
HIST1H1E	xclass I	2899171	-2.60	1.19×10^{-5}
HSD17B8	xclass II	2903488	-1.66	1.81×10^{-5}
MRPS18B	class I	2901818	2.02	1.81×10^{-5}
HLA-DPA1	class II	2950329	-1.29	2.18×10^{-5}
HIST1H2AC	xclass I	2899152	-1.87	2.23×10^{-5}
C6orf25	class III	2902609	-2.44	2.65×10^{-5}
EHMT2	class III	2949524	-1.49	2.73×10^{-5}
LOC554223	class I	2901043	1.62	3.06×10^{-5}
WDR46	xclass II	2950561	1.80	3.71×10^{-5}
HIST1H3E	xclass I	2899233	2.09	5.77×10^{-5}
HLA-DPB2	class II	2903435	-1.66	6.54×10^{-5}
PPP1R11	class I	2901352	1.61	8.30×10^{-5}
HIST2H4B	xclass I	2899216	-2.12	9.35×10^{-5}
HIST1H2AE	xclass I	2899223	-2.42	1.02×10^{-4}
AIF1	class III	2902444	-3.08	1.15×10^{-4}
BRD2	class II	2903343	1.34	1.42×10^{-4}
HIST1H2BN	xclass I	2900074	-1.70	2.20×10^{-4}
TUBB	class I	2948587	-1.42	2.59×10^{-4}
TRIM38	xclass I	2899022	1.41	2.82×10^{-4}
SLC39A7	xclass II	2903470	1.38	3.45×10^{-4}
BTN3A3	xclass I	2899413	1.66	3.95×10^{-4}
PFDN6	xclass II	2903588	1.51	3.96×10^{-4}
PPP1R18	class I	2948522	-1.78	5.64×10^{-4}
LST1	class III	2902427	-1.66	5.75×10^{-4}
TRIM27	xclass I	2947572	1.32	5.97×10^{-4}
PGBD1	xclass I	2900453	-1.27	9.05×10^{-4}
ATAT1	class I	2901841	-1.20	9.21×10^{-4}
HLA-E	class I	2901620	1.27	1.04×10^{-3}
HIST2H4B	xclass I	2899768	1.93	1.24×10^{-3}
BTN2A2	xclass I	2899340	-1.82	1.28×10^{-3}
HCG4	class I	2948024	1.75	1.28×10^{-3}

Gene	Region	Transcript cluster ID	Fold change	P_{FDR}
ZNF187	xclass I	2900423	-1.44	1.60×10^{-3}
C6orf10	class II	2949971	-1.75	1.72×10^{-3}
DOM3Z	class III	2949588	-1.31	1.84×10^{-3}
TRIM10	class I	2948239	-1.30	1.96×10^{-3}
ABHD16A	class III	2949256	1.24	2.79×10^{-3}
HIST1H3I	xclass I	2947077	-1.98	3.44×10^{-3}
NEU1	class III	2949471	1.44	3.96×10^{-3}
PRR3	class I	2901660	1.40	4.33×10^{-3}
HIST1H1B	xclass I	2947073	-2.32	4.64×10^{-3}
ZNF323	xclass I	2947248	-1.18	5.25×10^{-3}
HSPA1A	class I	2948713	1.26	5.45×10^{-3}
ABT1	xclass I	2899519	1.25	5.86×10^{-3}
ZSCAN16	xclass I	2900269	1.27	6.57×10^{-3}
HCG9	class I	2901312	1.49	7.66×10^{-3}
SLC17A4	xclass I	2898986	-1.46	7.74×10^{-3}
LOC100507547	class III	2949772	1.34	8.71×10^{-3}
HLA-DRA	class II	2903189	-1.14	9.89×10^{-3}
HCG27	class I	2902207	1.38	1.11×10^{-2}
HIST1H2BO	xclass I	2900116	-1.57	1.17×10^{-2}
MIR4640	class I	2901970	-1.12	1.20×10^{-2}
RXRB	xclass II	2950474	1.17	1.56×10^{-2}
ZKSCAN3	xclass I	2900497	-1.11	1.70×10^{-2}
HIST1H2AM	xclass I	2947100	-1.24	1.94×10^{-2}
VARS	class III	2949380	-1.21	1.94×10^{-2}
HIST1H4F	xclass I	2899243	1.76	2.04×10^{-2}
GPSM3	class III	2949885	-1.22	2.23×10^{-2}
STK19	class III	2902935	1.32	2.43×10^{-2}
APOM	class III	2902531	1.09	2.55×10^{-2}
BTN3A1	xclass I	2899372	-1.20	2.57×10^{-2}
HSPA1L	class III	2949450	1.28	2.66×10^{-2}
GPX6	xclass I	2947387	-1.32	2.70×10^{-2}
C2	class III	2902804	-1.22	3.00×10^{-2}
COL11A2	xclass II	2950384	-1.11	3.24×10^{-2}
MAS1L	xclass I	2947842	-1.20	3.29×10^{-2}
ZBTB22	xclass II	2950656	-1.18	3.43×10^{-2}

Gene	Region	Transcript cluster ID	Fold change	P_{FDR}
HIST1H2AA	xclass I	2946050	-1.32	3.46×10^{-2}
HIST1H2BM	xclass I	2900059	-1.48	3.54×10^{-2}
HIST1H3C	xclass I	2899102	-1.20	3.56×10^{-2}
PPP1R11	class I	2948188	-1.11	4.03×10^{-2}
SLC17A1	xclass I	2946056	-1.15	4.12×10^{-2}
SLC17A3	xclass I	2946106	-1.20	4.18×10^{-2}
HIST1H2BG	xclass I	2946345	-1.33	4.72×10^{-2}
BTN3A2	xclass I	2899298	1.38	4.76×10^{-2}
C6orf48	class III	2902736	-1.12	4.89×10^{-2}

All genes were differentially expressed at an FDR < 5%.

Table F2 The effect of IFN-α stimulation on alternative splicing in xMHC genes in B cells.

Gene	Region	Transcript cluster ID	P_{FDR}
Strong evidence of alternative splicing (FDR < 1%)			
HIST1H2AH	xclass I	2899756	3.26×10^{-9}
HIST1H2BD	xclass I	2899176	8.71×10^{-8}
HIST1H2AM	xclass I	2947100	1.89×10^{-6}
SKIV2L	class III	2902884	2.08×10^{-6}
HIST1H3D	xclass I	2946324	3.78×10^{-6}
NOTCH4	class III	2949901	4.77×10^{-6}
HCP5	class I	2902326	2.37×10^{-4}
PPP1R11	class I	2948188	2.52×10^{-4}
SCAND3	xclass I	2947405	6.08×10^{-4}
OR11A1	xclass I	2947805	6.59×10^{-4}
CSNK2B	class III	2902559	1.85×10^{-3}
CYP21A1P	class III	2903034	3.28×10^{-3}
MDC1	class I	2948564	4.70×10^{-3}
VWA7	class III	2949352	8.07×10^{-3}
Suggestive evidence of alternative splicing (FDR < 5%)			
COL11A2	xclass II	2950384	1.27×10^{-2}
C6orf47	class III	2949197	1.34×10^{-2}
GTF2H4	class I	2902013	2.15×10^{-2}
ZNF193	xclass I	2900372	3.01×10^{-2}
HIST1H2BJ	xclass I	2946681	3.52×10^{-2}
HIST1H3F	xclass I	2946364	4.27×10^{-2}

Appendix G

Table G1 HLA allelic composition in *DRB1*03:01*-homozygous and *DRB1*03:01*-heterozygous twins participating in the Affymetrix exon array study.

Twin ID	HLA-A allele 1	HLA-A allele 2	HLA-C allele 1	HLA-C allele 2	HLA-B allele 1	HLA-B allele 2	HLA-DRB1 allele 1	HLA-DRB1 allele 2	HLA-DQB1 allele 1	HLA-DQB1 allele 2	HLA-DPB1 allele 1	HLA-DPB1 allele 2
2151	A_0101	A_0101	C_0701	C_0701	B_0801	B_0801	DRB_0301	DRB_0301	DQB_0201	DQB_0201	-	-
2152	-	-	-	-	-	-	-	-	-	-	-	-
3571	A_0301	A_2301	C_0501	C_0701	B_0801	B_1801	DRB_0301	DRB_0301	DQB_0201	DQB_0201	DPB1_0301	DPB1_0101/ DPB1_0502
3572	A_0101	A_2301	C_0501	C_0701	B_0801	B_1801	DRB_0301	DRB_0301	DQB_0201	DQB_0201	DPB1_0101	DPB1_0301
30301	A_0101	A_0301	C_0401	C_0701	B_0801	B_3501	DRB_0301	DRB_0301	DQB_0201	DQB_0201	DPB1_0301	DPB1_0401
30302	-	-	-	-	-	-	-	-	-	-	-	-
32331	A_0101	A_2902	C_0501	C_0701	B_0801	B_4402	DRB_0101/ DRB_0103	DRB_0301	DQB_0201	DQB_0501	DPB1_0401	DPB1_0101
34572	-	-	-	-	-	-	-	-	-	-	-	-
45201	-	-	-	-	-	-	-	-	-	-	-	-
45202	-	-	-	-	-	-	-	-	-	-	-	-
51351	-	-	-	-	-	-	-	-	-	-	-	-
51352	A_0101	A_0101	C_0701	C_0701	B_0801	B_0801	DRB_0301	DRB_0301	DQB_0201	DQB_0201	DPB1_0401	DPB1_0401
64522	A_0101	A_0101	C_0701	C_0701	B_0801	B_0801	DRB_0301	DRB_0301	DQB_0201	DQB_0201	DPB1_1101	DPB1_0101
73511	A_0101	A_0101	C_0701	C_0701	B_0801	B_0801	DRB_0301	DRB_0301	DQB_0201	DQB_0201	DPB1_0301	DPB1_0301
73512	A_0101	A_0101	C_0701	C_0701	B_0801	B_0801	DRB_0301	DRB_0301	DQB_0201	DQB_0201	DPB1_0301	DPB1_0301
75591	A_0101	A_6801	C_0701	C_0701	B_0801	B_0801	DRB_0301	DRB_0301	DQB_0201	DQB_0201	-	-
75592	A_0101	A_6801	C_0701	C_0701	B_0801	B_0801	DRB_0301	DRB_0301	DQB_0201	DQB_0201	-	-
81231	-	-	-	-	-	-	-	-	-	-	-	-
95711	A_0101	A_0101	C_0701	C_0701	B_0801	B_0801	DRB_0301	DRB_0301	DQB_0201	DQB_0201	DPB1_0401	DPB1_0401
95712	A_0101	A_0101	C_0701	C_0701	B_0801	B_0801	DRB_0301	DRB_0301	DQB_0201	DQB_0201	DPB1_0401	DPB1_0401
98661	A_0101	A_0201	C_0501	C_0701	B_0801	B_4402	DRB_0301	DRB_0301	DQB_0201	DQB_0201	-	-
98662	-	-	-	-	-	-	-	-	-	-	-	-
COX	**A_0101**	**A_0101**	**C_0701**	**C_0701**	**B_0801**	**B_0801**	**DRB_0301**	**DRB_0301**	**DQB_0201**	**DQB_0201**	**DPB1_0301**	**DPB1_0301**
QBL	**A_2601**	**A_2601**	**C_0501**	**C_0501**	**B_1801**	**B_1801**	**DRB_0301**	**DRB_0301**	**DQB_0201**	**DQB_0201**	**DPB1_0202**	**DPB1_0202**

HLA imputation was conducted by Dr David Morris (King's College London), using the IMPUTE tool. Dashes indicate that imputed data was not available for that individual.

Appendix H

```
BTN3A1      ATGAAAATGGCAAGTTTCCTGGCCTTCCTTCTGCTCAACTTTCGTGTCTGCCTCCTTTTG
BTN3A2      ATGAAAATGGCAAGTTCCCTGGCTTTCCTTCTGCTCAACTTTCATGTCTCCCTCCTCTTG
BTN3A3      ATGAAAATGGCAAGTTCCCTGGCTTTCCTTCTGCTCAACTTTCATGTCTCCCTCTTCTTG
            ********************************************************** *** 

BTN3A1      CTTCAGCTGCTCATGCCTCACTCAGCTCAGTTTTCTGTGCTTGGACCCTCTGGGCCCATC
BTN3A2      GTCCAGCTGCTCACTCCTTGCTCAGCTCAGTTTTCTGTGCTTGGACCCTCTGGGCCCATC
BTN3A3      GTCCAGCTGCTCACTCCTTGCTCAGCTCAGTTTTCTGTGCTTGGACCCTCTGGGCCCATC
             * **********  ***  ************************************

BTN3A1      CTGGCCATGGTGGGTGAAGACGCTGATCTGCCCTGTCACCTGTTCCCGACCATGAGTGCA
BTN3A2      CTGGCCATGGTGGGTGAAGACGCTGATCTGCCCTGTCACCTGTTCCCGACCATGAGTGCA
BTN3A3      CTGGCCATGGTGGGTGAAGACGCTGATCTGCCCTGTCACCTGTTCCCGACCATGAGTGCA
            ************************************************************

BTN3A1      GAGACCATGGAGCTGAAGTGGGTGAGTTCCAGCCTAAGGCAGGTGGTGAACGTGTATGCA
BTN3A2      GAGACCATGGAGCTGAAGTGGGTAAGTTCCAGCCTAAGGCAGGTGGTGAACGTGTATGCA
BTN3A3      GAGACCATGGAGCTGAGGTGGGTGAGTTCCAGCCTAAGGCAGGTGGTGAACGTGTATGCA
            ********************************************************** 

BTN3A1      GATGGAAAGGAAGTGGAAGACAGGCAGAGTGCACCGTATCGAGGGAGAACTTCGATTCTG
BTN3A2      GATGGAAAGGAAGTGGAAGACAGGCAGAGTGCACCGTATCGAGGGAGAACTTCGATTCTG
BTN3A3      GATGGAAAGGAAGTGGAAGACAGGCAGAGTGCACCGTATCGAGGGAGAACTTCGATTCTG
            ************************************************************

BTN3A1      CGGGATGGCATCACTGCAGGGAAGGCTGCTCTCCGAATACACAACGTCACAGCCTCTGAC
BTN3A2      CGGGATGGCATCACTGCAGGGAAGGCTGCTCTCCGAATACACAACGTCACAGCCTCTGAC
BTN3A3      CGGGATGGCATCACTGCAGGGAAGGCTGCTCTCCGAATACACAACGTCACAGCCTCTGAC
            ************************************************************

BTN3A1      AGTGGAAAGTACTTGTGTTATTTCCAAGATGGTGACTTCTATGAAAAAGCCCTGGTGGAG
BTN3A2      AGTGGAAAGTACTTGTGTTATTTCCAAGATGGTGACTTCTATGAAAAAGCCCTGGTGGAG
BTN3A3      AGTGGAAAGTACTTGTGTTATTTCCAAGATGGTGACTTCTACGAAAAAGCCCTGGTGGAG
            ************************************** ******************

BTN3A1      CTGAAGGTTGCAGCACTGGGTTCTGATCTTCACGTTGATGTGAAGGGTTACAAGGATGGA
BTN3A2      CTGAAGGTTGCAGCACTGGGTTCTAATCTTCACGTCGAAGTGAAGGGTTATGAGGATGGA
BTN3A3      CTGAAGGTTGCAGCATTGGGTTCTGATCTTCACATTGAAGTGAAGGGTTATGAGGATGGA
            *************** ******** ******** * ** ************  ********

BTN3A1      GGGATCCATCTGGAGTGCAGGTCCACTGGCTGGTACCCCCAACCCCAAATACAGTGGAGC
BTN3A2      GGGATCCATCTGGAGTGCAGGTCCACCGGCTGGTACCCCCAACCCCAAATACAGTGGAGC
BTN3A3      GGGATCCATCTGGAGTGCAGGTCCACTGGCTGGTACCCCCAACCCCAAATAAAGTGGAGC
            ************************** ************************ ********

BTN3A1      AACAACAAGGGAGAGAACATCCCGACTGTGGAAGCACCTGTGGTTGCAGACGGAGTGGGC
BTN3A2      AACGCCAAGGGAGAGAACATCCCAGCTGTGGAAGCACCTGTGGTTGCAGATGGAGTGGGC
BTN3A3      GACACCAAGGGAGAGAACATCCCGGCTGTGGAAGCACCTGTGGTTGCAGATGGAGTGGGC
             **  ******************  ************************* *********

BTN3A1      CTGTATGCAGTAGCAGCATCTGTGATCATGAGAGGCAGCTCTGGGGAGGGTGTATCCTGT
BTN3A2      CTATATGAAGTAGCAGCATCTGTGATCATGAGAGGCGGCTCCGGGGAGGGTGTATCCTGC
BTN3A3      CTGTATGCAGTAGCAGCATCTGTGATCATGAGAGGCAGCTCTGGTGGGGGTGTATCCTGC
            ** **** ***************************** **** ** * ************ 

BTN3A1      ACCATCAGAAGTTCCCTCCTCGGCCTGGAAAAGACAGCCAGCATTTCCATCGCAGACCCC
BTN3A2      ATCATCAGAAATTCCCTCCTCGGCCTGGAAAAGACAGCCAGCATTTCCATCGCAGACCCC
BTN3A3      ATCATCAGAAATTCCCTCCTCGGCCTGGAAAAGACAGCCAGCATATCCATCGCAGACCCC
            * ******** ********************************** ***************

BTN3A1      TTCTTCAGGAGCGCCCAGAGGTGGATCGCCGCCCTGGCAGGGACCCTGCCTGTCTTGCTG
BTN3A2      TTCTTCAGGAGCGCCCAGCCCTGGATCGCAGCCCTGGCAGGGACCCTGCCTATCTTGCTG
BTN3A3      TTCTTCAGGAGCGCCCAGCCCTGGATCGCGGCCCTGGCAGGGACCCTGCCTATCTCGTTG
            ******************   ******** ********************* *** * **
```

```
BTN3A1      CTGCTTCTTGGGGGAGCCGGTTACTTCCTGTGGCAACAGCAGGAGGAAAAAAAGACTCAG
BTN3A2      CTGCTTCTCGCCGGAGCCAGTTACTTCTTGTGGAGACAACAGAAGGAAATAACTGCTCTG
BTN3A3      CTGCTTCTCGCAGGAGCCAGTTACTTCTTGTGGAGACAACAGAAGGAAAAAATTGCTCTG
            ******** *  ****** ******** *****  *** *** ****** **   *** *

BTN3A1      TTCAGAAAGAAAAAGAGAGAGCAAGAGTTGAGAGAAATGGCATGGAGCACAATGAAGCAA
BTN3A2      TCCAGTGAGATAGAAAGTGAGCAAGAGATGAAAGAAATGGGATATGCTGCAACAGAGCGG
BTN3A3      TCCAGGGAGACAGAAAGAGAGCGAGAGATGAAAGAAATGGGATACGCTGCAACAGAGCAA
            * ***  *** * * ** **** **** *** ******** **      ***   ***

BTN3A1      GAACAAAGCACAAGAGTGAAGCTCCTGGAGGAACTCAGATGGAGAAGTATCCAGTATGCA
BTN3A2      GAAATAAGCCTAAGAGAGAGCCTCCAGGAGGAACTCAAGAGGAAAAAAATCCAGTACTTG
BTN3A3      GAAATAAGCCTAAGAGAGAAGCTCCAGGAGGAACTCAAGTGGAGGAAAATCCAGTACATG
            ***  ****  ***** **  **** **********   ***  *  ********

BTN3A1      TCTCGGGGAGAGAGACATTCAGCCTATAATGAATGGAAAAAGGCCCTCTTCAAGCCTGCG
BTN3A2      ACTCGTGGAGAGGAGTCTTCGTCCGATACCAATAAGTCAGCCTGA---------------
BTN3A3      GCTCGTGGAGAGAAGTCTTTGGCCTATCATGAATGGAAAATGGCCCTCTTCAAACCTGCG
             **** ******     **   ** **    *   *  *

BTN3A1      GATGTGATTCTGGATCCAAAAACAGCAAACCCCATCCTCCTTGTTTCTGAGGACCAGAGG
BTN3A2      ------------------------------------------------------------
BTN3A3      GATGTGATTCTGGATCCAGACACGGCAAACGCCATCCTCCTTGTTTCTGAGGACCAGAGG

BTN3A1      AGTGTGCAGCGTGCCAAGGAGCCCCAGGATCTGCCAGACAACCCTGAGAGATTTAATTGG
BTN3A2      ------------------------------------------------------------
BTN3A3      AGTGTGCAGCGTGCTGAAGAGCCGCGGGATCTGCCAGACAACCCTGAGAGATTTGAATGG

BTN3A1      CATTATTGTGTTCTCGGCTGTGAGAGCTTCATATCAGGGAGACATTACTGGGAGGTGGAG
BTN3A2      ------------------------------------------------------------
BTN3A3      CGTTACTGTGTCCTTGGCTGTGAAAACTTCACATCAGGGAGACATTACTGGGAGGTGGAA

BTN3A1      GTAGGGGACAGGAAAGAGTGGCATATAGGGGTGTGCAGTAAGAATGTGCAGA---GAAAA
BTN3A2      ------------------------------------------------------------
BTN3A3      GTGGGGGACAGAAAAGAGTGGCATATTGGGGTATGTAGTAAGAACGTGGAGAGGAAAAAA

BTN3A1      GGCTGGGTCAAAATGACACCTGAGAATGGATTCTGGACTATGGGGCTGACTGATGGGAAT
BTN3A2      ------------------------------------------------------------
BTN3A3      GGTTGGGTCAAAATGACACCGGAGAACGGATACTGGACTATGGGCCTGACTGATGGGAAT

BTN3A1      AAGTATCGGACTCTAACTGAGCCCAGAACCAACCTGAAACTTCCTAAGCCCCCTAAGAAA
BTN3A2      ------------------------------------------------------------
BTN3A3      AAGTATCGGGCTCTCACTGAGCCCAGAACCAACCTGAAACTTCCTGAGCCTCCTAGGAAA

BTN3A1      GTGGGGGTCTTCCTGGACTATGAGACTGGAGATATCTCATTCTACAATGCTGTGGATGGA
BTN3A2      ------------------------------------------------------------
BTN3A3      GTGGGGATCTTCCTGGACTATGAGACTGGAGAGATCTCGTTCTATAATGCCACAGATGGA

BTN3A1      TCGCATATTCATACTTTCCTGGACGTCTCCTTCTCTGAGGCTCTATATCCTGTTTTCAGA
BTN3A2      ------------------------------------------------------------
BTN3A3      TCTCATATCTACACCTTTCCGCACGCCTCTTTCTCTGAGCCTCTATATCCTGTTTTCAGA

BTN3A1      ATTTTGACCTTGGAGCCCACGGCCCTGACTATTTGTCCAGCGTGA---------------
BTN3A2      ------------------------------------------------------------
BTN3A3      ATTTTGACCTTGGAGCCCACTGCCCTGACCATTTGCCCAATACCAAAAGAAGTAGAGAGT

BTN3A1      ------------------------------------------------------------
BTN3A2      ------------------------------------------------------------
BTN3A3      TCCCCCGATCCTGACCTAGTGCCTGATCATTCCCTGGAGACACCACTGACCCCGGGCTTA
```

Location of the TaqMan qPCR amplicon

```
BTN3A1      ------------------------------------------------------------
BTN3A2      ------------------------------------------------------------
BTN3A3      GCTAATGAAAGTGGGGAGCCTCAGGCTGAAGTAACATCTCTGCTTCTCCCTGCCCACCCT


BTN3A1      ------------------------------------------------------------
BTN3A2      ------------------------------------------------------------
BTN3A3      GGAGCTGAGGTCTCCCCTTCTGCAACAACCAATCAGAACCATAAGCTACAGGCACGCACT


BTN3A1      ---------------
BTN3A2      ---------------
BTN3A3      GAAGCACTTTACTGA
```

Figure H1 Clustal alignment of the coding cDNA sequences from all three BTN3 genes.

The *BTN3A2* exons targeted by the TaqMan qPCR assay are highlighted in yellow (exon 8) and cyan (exon 9). The asterisks denote sequence homology between *BTN3A1*, *BTN3A2*, and *BTN3A3*.

Appendix I

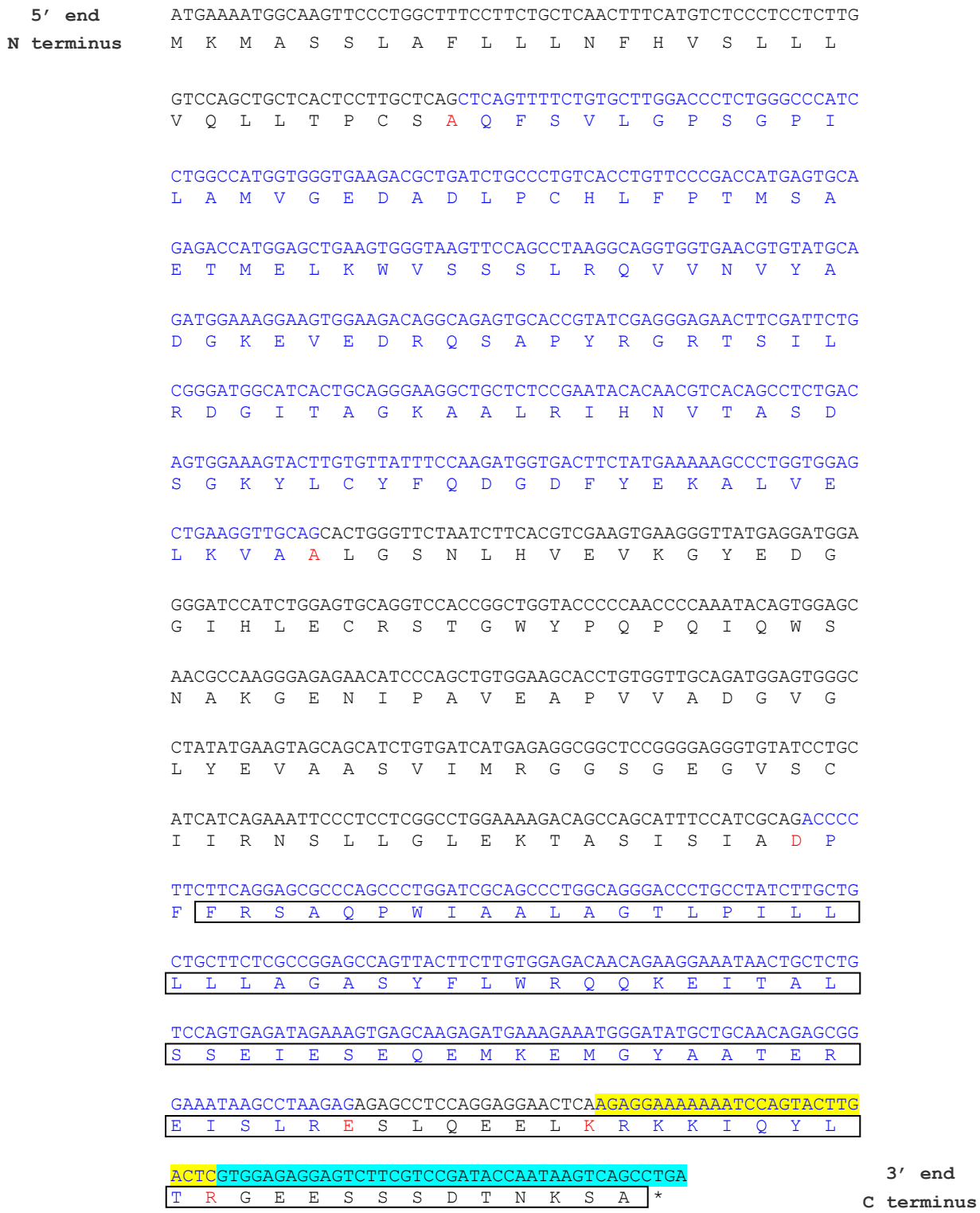


Figure I1 The *BTN3A2* coding cDNA sequence (top), and the corresponding BT3.2 amino acid sequence (bottom).

Alternating exons are coloured in black and blue, and residues over splice sites are shown in red. The boxed region indicates the antigen targeted by the BT3.2 antibody used for Western blotting. The corresponding cDNA sequence overlaps the exons targeted by the TaqMan qPCR assay, which are highlighted in yellow (exon 8) and cyan (exon 9). The asterisk denotes the stop codon.

Bibliography

ABDOU, A. M., GAO, X., COZEN, W., CERHAN, J. R., ROTHMAN, N., MARTIN, M. P., DAVIS, S., SCHENK, M., CHANOCK, S. J., HARTGE, P., CARRINGTON, M. & WANG, S. S. 2010. Human leukocyte antigen (HLA) A1-B8-DR3 (8.1) haplotype, tumor necrosis factor (TNF) G-308A, and risk of non-Hodgkin lymphoma. *Leukemia,* 24**,** 1055-8.

ABELER-DORNER, L., SWAMY, M., WILLIAMS, G., HAYDAY, A. C. & BAS, A. 2012. Butyrophilins: an emerging family of immune regulators. *Trends Immunol,* 33**,** 34-41.

ABELSON, A. K., DELGADO-VEGA, A. M., KOZYREV, S. V., SANCHEZ, E., VELAZQUEZ-CRUZ, R., ERIKSSON, N., WOJCIK, J., LINGA REDDY, M. V., LIMA, G., D'ALFONSO, S., MIGLIARESI, S., BACA, V., OROZCO, L., WITTE, T., ORTEGO-CENTENO, N., GROUP, A., ABDERRAHIM, H., PONS-ESTEL, B. A., GUTIERREZ, C., SUAREZ, A., GONZALEZ-ESCRIBANO, M. F., MARTIN, J. & ALARCON-RIQUELME, M. E. 2009. STAT4 associates with systemic lupus erythematosus through two independent effects that correlate with gene expression and act additively with IRF5 to increase risk. *Ann Rheum Dis,* 68**,** 1746-53.

ABRAHAM, L. J. & KROEGER, K. M. 1999. Impact of the -308 TNF promoter polymorphism on the transcriptional regulation of the TNF gene: relevance to disease. *J Leukoc Biol,* 66**,** 562-6.

ABSHER, D. M., LI, X., WAITE, L. L., GIBSON, A., ROBERTS, K., EDBERG, J., CHATHAM, W. W. & KIMBERLY, R. P. 2013. Genome-wide DNA methylation analysis of systemic lupus erythematosus reveals persistent hypomethylation of interferon genes and compositional changes to CD4+ T-cell populations. *PLoS Genet,* 9**,** e1003678.

AFSHAR, M., MARTINEZ, A. D., GALLO, R. L. & HATA, T. R. 2013. Induction and exacerbation of psoriasis with Interferon-alpha therapy for hepatitis C: a review and analysis of 36 cases. *J Eur Acad Dermatol Venereol,* 27**,** 771-8.

AL-MAYOUF, S. M., SUNKER, A., ABDWANI, R., ABRAWI, S. A., ALMURSHEDI, F., ALHASHMI, N., AL SONBUL, A., SEWAIRI, W., QARI, A., ABDALLAH, E., AL-OWAIN, M., AL MOTYWEE, S., AL-RAYES, H., HASHEM, M., KHALAK, H., AL-JEBALI, L. & ALKURAYA, F. S. 2011. Loss-of-function variant in DNASE1L3 causes a familial form of systemic lupus erythematosus. *Nat Genet,* 43**,** 1186-8.

ALARCON-RIQUELME, M. E., ZIEGLER, J. T., MOLINEROS, J., HOWARD, T. D., MORENO-ESTRADA, A., SANCHEZ-RODRIGUEZ, E., AINSWORTH, H. C., ORTIZ-TELLO, P., COMEAU, M. E., RASMUSSEN, A., KELLY, J. A., ADLER, A., ACEVEDO-VAZQUEZ, E. M., MARIANO CUCHO-VENEGAS, J., GARCIA-DE LA TORRE, I., CARDIEL, M. H., MIRANDA, P., CATOGGIO, L. J., MARADIAGA-CECENA, M., GAFFNEY, P. M., VYSE, T. J., CRISWELL, L. A., TSAO, B. P., SIVILS, K. L., BAE, S. C., JAMES, J. A., KIMBERLY, R. P., KAUFMAN, K. M., HARLEY, J. B., ESQUIVEL-VALERIO, J. A., MOCTEZUMA, J. F., GARCIA, M. A., BERBOTTO, G. A., BABINI, A. M., SCHERBARTH, H., TOLOZA, S., BACA, V., NATH, S. K., AGUILAR SALINAS, C., OROZCO, L., TUSIE-LUNA, T., ZIDOVETZKI, R., PONS-ESTEL, B. A., LANGEFELD, C. D. & JACOB, C. O. 2016. Genome-Wide Association Study in an Amerindian Ancestry Population Reveals Novel Systemic Lupus Erythematosus Risk Loci and the Role of European Admixture. *Arthritis Rheumatol,* 68**,** 932-43.

ALARCON-SEGOVIA, D., ALARCON-RIQUELME, M. E., CARDIEL, M. H., CAEIRO, F., MASSARDO, L., VILLA, A. R., PONS-ESTEL, B. A. & GRUPO LATINOAMERICANO DE ESTUDIO DEL LUPUS, E. 2005. Familial aggregation of systemic lupus erythematosus, rheumatoid arthritis, and other autoimmune diseases in 1,177 lupus patients from the GLADEL cohort. *Arthritis Rheum,* 52**,** 1138-47.

ALBERT, F. W. & KRUGLYAK, L. 2015. The role of regulatory variation in complex traits and disease. *Nat Rev Genet,* 16**,** 197-212.

AMISSAH-ARTHUR, M. B. & GORDON, C. 2010. Contemporary treatment of systemic lupus erythematosus: an update for clinicians. *Ther Adv Chronic Dis,* 1**,** 163-75.

APELBAUM, A., YARDEN, G., WARSZAWSKI, S., HARARI, D. & SCHREIBER, G. 2013. Type I interferons induce apoptosis by balancing cFLIP and caspase-8 independent of death ligands. *Mol Cell Biol,* 33**,** 800-14.

ARINGER, M., HOUSSIAU, F., GORDON, C., GRANINGER, W. B., VOLL, R. E., RATH, E., STEINER, G. & SMOLEN, J. S. 2009. Adverse events and efficacy of TNF-alpha

blockade with infliximab in patients with systemic lupus erythematosus: long-term follow-up of 13 patients. *Rheumatology (Oxford),* 48**,** 1451-4.
ARINGER, M. & SMOLEN, J. S. 2008. The role of tumor necrosis factor-alpha in systemic lupus erythematosus. *Arthritis Res Ther,* 10**,** 202.
ARNETT, H. A., ESCOBAR, S. S. & VINEY, J. L. 2009. Regulation of costimulation in the era of butyrophilins. *Cytokine,* 46**,** 370-5.
ASSASSI, S., MAYES, M. D., ARNETT, F. C., GOURH, P., AGARWAL, S. K., MCNEARNEY, T. A., CHAUSSABEL, D., OOMMEN, N., FISCHBACH, M., SHAH, K. R., CHARLES, J., PASCUAL, V., REVEILLE, J. D. & TAN, F. K. 2010. Systemic sclerosis and lupus: points in an interferon-mediated continuum. *Arthritis Rheum,* 62**,** 589-98.
AYED, K., GORGI, Y., AYED-JENDOUBI, S. & BARDI, R. 2004. The involvement of HLA - DRB1*, DQA1*, DQB1* and complement C4A loci in diagnosing systemic lupus erythematosus among Tunisians. *Ann Saudi Med,* 24**,** 31-5.
BAECHLER, E. C., BATLIWALLA, F. M., KARYPIS, G., GAFFNEY, P. M., ORTMANN, W. A., ESPE, K. J., SHARK, K. B., GRANDE, W. J., HUGHES, K. M., KAPUR, V., GREGERSEN, P. K. & BEHRENS, T. W. 2003. Interferon-inducible gene expression signature in peripheral blood cells of patients with severe lupus. *Proc Natl Acad Sci U S A,* 100**,** 2610-5.
BANG, S. Y., CHOI, J. Y., PARK, S., CHOI, J., HONG, S. J., LEE, H. S., CHOI, C. B. & BAE, S. C. 2015. Influence of Susceptibility HLA-DRB1 alleles on the clinical subphenotypes of Systemic Lupus Erythematosus in Koreans. *Arthritis Rheumatol*.
BANNWARTH, S., FIGUEROA, A., FRAGAKI, K., DESTROISMAISONS, L., LACAS-GERVAIS, S., LESPINASSE, F., VANDENBOS, F., PRADELLI, L.A., RICCI, J., RÖTIG, A., MICHIELS, J., VELDE, C.V., PAQUIS-FLUCKLINGER, V. 2012. The human MSH5 (MutSHomolog 5) protein localizes to mitochondria and protects the mitochondrial genome from oxidative damage. *Mitochondrion*, 12, 654–665.
BARCELLOS, L. F., MAY, S. L., RAMSAY, P. P., QUACH, H. L., LANE, J. A., NITITHAM, J., NOBLE, J. A., TAYLOR, K. E., QUACH, D. L., CHUNG, S. A., KELLY, J. A., MOSER, K. L., BEHRENS, T. W., SELDIN, M. F., THOMSON, G., HARLEY, J. B., GAFFNEY, P. M. & CRISWELL, L. A. 2009. High-density SNP screening of the major histocompatibility complex in systemic lupus erythematosus demonstrates strong evidence for independent susceptibility regions. *PLoS Genet,* 5**,** e1000696.
BAUER, J. W., BAECHLER, E. C., PETRI, M., BATLIWALLA, F. M., CRAWFORD, D., ORTMANN, W. A., ESPE, K. J., LI, W., PATEL, D. D., GREGERSEN, P. K. & BEHRENS, T. W. 2006. Elevated serum levels of interferon-regulated chemokines are biomarkers for active human systemic lupus erythematosus. *PLoS Med,* 3**,** e491.
BECKER, A. M., DAO, K. H., HAN, B. K., KORNU, R., LAKHANPAL, S., MOBLEY, A. B., LI, Q. Z., LIAN, Y., WU, T., REIMOLD, A. M., OLSEN, N. J., KARP, D. R., CHOWDHURY, F. Z., FARRAR, J. D., SATTERTHWAITE, A. B., MOHAN, C., LIPSKY, P. E., WAKELAND, E. K. & DAVIS, L. S. 2013. SLE peripheral blood B cell, T cell and myeloid cell transcriptomes display unique profiles and each subset contributes to the interferon signature. *PLoS One,* 8**,** e67003.
BELOT, A. & CIMAZ, R. 2012. Monogenic forms of systemic lupus erythematosus: new insights into SLE pathogenesis. *Pediatr Rheumatol Online J,* 10**,** 21.
BELOT, A., KASHER, P. R., TROTTER, E. W., FORAY, A. P., DEBAUD, A. L., RICE, G. I., SZYNKIEWICZ, M., ZABOT, M. T., ROUVET, I., BHASKAR, S. S., DALY, S. B., DICKERSON, J. E., MAYER, J., O'SULLIVAN, J., JUILLARD, L., URQUHART, J. E., FAWDAR, S., MARUSIAK, A. A., STEPHENSON, N., WASZKOWYCZ, B., M, W. B., BIESECKER, L. G., G, C. M. B., RENE, C., ELIAOU, J. F., FABIEN, N., RANCHIN, B., COCHAT, P., GAFFNEY, P. M., ROZENBERG, F., LEBON, P., MALCUS, C., CROW, Y. J., BROGNARD, J. & BONNEFOY, N. 2013. Protein kinase cdelta deficiency causes mendelian systemic lupus erythematosus with B cell-defective apoptosis and hyperproliferation. *Arthritis Rheum,* 65**,** 2161-71.
BENTHAM, J., MORRIS, D. L., CUNNINGHAME GRAHAM, D. S., PINDER, C. L., TOMBLESON, P., BEHRENS, T. W., MARTIN, J., FAIRFAX, B. P., KNIGHT, J. C., CHEN, L., REPLOGLE, J., SYVANEN, A. C., RONNBLOM, L., GRAHAM, R. R., WITHER, J. E., RIOUX, J. D., ALARCON-RIQUELME, M. E. & VYSE, T. J. 2015. Genetic association analyses implicate aberrant regulation of innate and adaptive immunity genes in the pathogenesis of systemic lupus erythematosus. *Nat Genet,* 47**,** 1457-64.
BERTHIER, C. C., KRETZLER, M. & DAVIDSON, A. 2012. From the Large Scale Expression Analysis of Lupus Nephritis to Targeted Molecular Medicine. *J Data Mining Genomics Proteomics,* 3.

BETTINOTTI, M. P., HARTUNG, K., DEICHER, H., MESSER, G., KELLER, E., WEISS, E. H. & ALBERT, E. D. 1993. Polymorphism of the tumor necrosis factor beta gene in systemic lupus erythematosus: TNFB-MHC haplotypes. *Immunogenetics,* 37**,** 449-54.

BISHOF, N. A., WELCH, T. R., BEISCHEL, L. S., CARSON, D. & DONNELLY, P. A. 1993. DP polymorphism in HLA-A1,-B8,-DR3 extended haplotypes associated with membranoproliferative glomerulonephritis and systemic lupus erythematosus. *Pediatr Nephrol,* 7**,** 243-6.

BOACKLE, S. A. 2013. Advances in lupus genetics. *Curr Opin Rheumatol,* 25**,** 561-8.

BOTEVA, L., MORRIS, D. L., CORTES-HERNANDEZ, J., MARTIN, J., VYSE, T. J. & FERNANDO, M. M. 2012. Genetically determined partial complement C4 deficiency states are not independent risk factors for SLE in UK and Spanish populations. *Am J Hum Genet,* 90**,** 445-56.

BOTTINI, N. & PETERSON, E. J. 2014. Tyrosine phosphatase PTPN22: multifunctional regulator of immune signaling, development, and disease. *Annu Rev Immunol,* 32**,** 83-119.

BOWIE, A. G. & UNTERHOLZNER, L. 2008. Viral evasion and subversion of pattern-recognition receptor signalling. *Nat Rev Immunol,* 8**,** 911-22.

BRAUN, D., CARAMALHO, I. & DEMENGEOT, J. 2002. IFN-alpha/beta enhances BCR-dependent B cell responses. *Int Immunol,* 14**,** 411-9.

BREEN, E. C., HUSSAIN, S. K., MAGPANTAY, L., JACOBSON, L. P., DETELS, R., RABKIN, C. S., KASLOW, R. A., VARIAKOJIS, D., BREAM, J. H., RINALDO, C. R., AMBINDER, R. F. & MARTINEZ-MAZA, O. 2011. B-cell stimulatory cytokines and markers of immune activation are elevated several years prior to the diagnosis of systemic AIDS-associated non-Hodgkin B-cell lymphoma. *Cancer Epidemiol Biomarkers Prev,* 20**,** 1303-14.

BRIGGS, T. A., RICE, G. I., DALY, S., URQUHART, J., GORNALL, H., BADER-MEUNIER, B., BASKAR, K., BASKAR, S., BAUDOUIN, V., BERESFORD, M. W., BLACK, G. C., DEARMAN, R. J., DE ZEGHER, F., FOSTER, E. S., FRANCES, C., HAYMAN, A. R., HILTON, E., JOB-DESLANDRE, C., KULKARNI, M. L., LE MERRER, M., LINGLART, A., LOVELL, S. C., MAURER, K., MUSSET, L., NAVARRO, V., PICARD, C., PUEL, A., RIEUX-LAUCAT, F., ROIFMAN, C. M., SCHOLL-BURGI, S., SMITH, N., SZYNKIEWICZ, M., WIEDEMAN, A., WOUTERS, C., ZEEF, L. A., CASANOVA, J. L., ELKON, K. B., JANCKILA, A., LEBON, P. & CROW, Y. J. 2011. Tartrate-resistant acid phosphatase deficiency causes a bone dysplasia with autoimmunity and a type I interferon expression signature. *Nat Genet,* 43**,** 127-31.

BROWN, E. E., EDBERG, J. C. & KIMBERLY, R. P. 2007. Fc receptor genes and the systemic lupus erythematosus diathesis. *Autoimmunity,* 40**,** 567-81.

BRYAN, A. R. & WU, E. Y. 2014. Complement deficiencies in systemic lupus erythematosus. *Curr Allergy Asthma Rep,* 14**,** 448.

BURN, G. L., SVENSSON, L., SANCHEZ-BLANCO, C., SAINI, M. & COPE, A. P. 2011. Why is PTPN22 a good candidate susceptibility gene for autoimmune disease? *FEBS Lett,* 585**,** 3689-98.

BUTTMANN, M. & RIECKMANN, P. 2007. Interferon-beta1b in multiple sclerosis. *Expert Rev Neurother,* 7**,** 227-39.

CHAM, C. M., KO, K. & NIEWOLD, T. B. 2012. Interferon regulatory factor 5 in the pathogenesis of systemic lupus erythematosus. *Clin Dev Immunol,* 2012**,** 780436.

CHANG, N. H., MCKENZIE, T., BONVENTI, G., LANDOLT-MARTICORENA, C., FORTIN, P. R., GLADMAN, D., UROWITZ, M. & WITHER, J. E. 2008. Expanded population of activated antigen-engaged cells within the naive B cell compartment of patients with systemic lupus erythematosus. *J Immunol,* 180**,** 1276-84.

CHRISTIAN, N., SMIKLE, M. F., DECEULAER, K., DANIELS, L., WALRAVENS, M. J. & BARTON, E. N. 2007. Antinuclear antibodies and HLA class II alleles in Jamaican patients with systemic lupus erythematosus. *West Indian Med J,* 56**,** 130-3.

CLANCY, R. M., MARION, M. C., KAUFMAN, K. M., RAMOS, P. S., ADLER, A., INTERNATIONAL CONSORTIUM ON SYSTEMIC LUPUS ERYTHEMATOSUS, G., HARLEY, J. B., LANGEFELD, C. D. & BUYON, J. P. 2010. Identification of candidate loci at 6p21 and 21q22 in a genome-wide association study of cardiac manifestations of neonatal lupus. *Arthritis Rheum,* 62**,** 3415-24.

CLARK, T. A., SCHWEITZER, A. C., CHEN, T. X., STAPLES, M. K., LU, G., WANG, H., WILLIAMS, A. & BLUME, J. E. 2007. Discovery of tissue-specific exons using comprehensive human exon microarrays. *Genome Biol,* 8**,** R64.

COICO, R. & SUNSHINE, G. 2015. *Immunology: A Short Course*, Wiley.

COLTEN, H. R., STRUNK, R. C., PERLMUTTER, D. H. & COLE, F. S. 1986. Regulation of complement protein biosynthesis in mononuclear phagocytes. *Ciba Found Symp,* 118**,** 141-54.

CORO, E. S., CHANG, W. L. & BAUMGARTH, N. 2006. Type I IFN receptor signals directly stimulate local B cells early following influenza virus infection. *J Immunol,* 176**,** 4343-51.

CORTES, L. M., BALTAZAR, L. M., LOPEZ-CARDONA, M. G., OLIVARES, N., RAMOS, C., SALAZAR, M., SANDOVAL, L., LORENZ, M. G., CHAKRABORTY, R., PATERSON, A. D. & RIVAS, F. 2004. HLA class II haplotypes in Mexican systemic lupus erythematosus patients. *Hum Immunol,* 65**,** 1469-76.

COSTA-REIS, P. & SULLIVAN, K. E. 2013. Genetics and epigenetics of systemic lupus erythematosus. *Curr Rheumatol Rep,* 15**,** 369.

CROW, M. K. 2010. Type I interferon in organ-targeted autoimmune and inflammatory diseases. *Arthritis Res Ther,* 12 Suppl 1**,** S5.

CROW, Y. J. & MANEL, N. 2015. Aicardi-Goutieres syndrome and the type I interferonopathies. *Nat Rev Immunol,* 15**,** 429-40.

CUBILLOS-RUIZ, J. R. & CONEJO-GARCIA, J. R. 2011. It never rains but it pours: potential role of butyrophilins in inhibiting anti-tumor immune responses. *Cell Cycle,* 10**,** 368-9.

CUNNINGHAME GRAHAM, D. S., GRAHAM, R. R., MANKU, H., WONG, A. K., WHITTAKER, J. C., GAFFNEY, P. M., MOSER, K. L., RIOUX, J. D., ALTSHULER, D., BEHRENS, T. W. & VYSE, T. J. 2008. Polymorphism at the TNF superfamily gene TNFSF4 confers susceptibility to systemic lupus erythematosus. *Nat Genet,* 40**,** 83-9.

DANG, J., SHAN, S., LI, J., ZHAO, H., XIN, Q., LIU, Y., BIAN, X. & LIU, Q. 2014. Gene-gene interactions of IRF5, STAT4, IKZF1 and ETS1 in systemic lupus erythematosus. *Tissue Antigens,* 83**,** 401-8.

DE GOER DE HERVE, M. G., DURALI, D., DEMBELE, B., GIULIANI, M., TRAN, T. A., AZZARONE, B., EID, P., TARDIEU, M., DELFRAISSY, J. F. & TAOUFIK, Y. 2011. Interferon-alpha triggers B cell effector 1 (Be1) commitment. *PLoS One,* 6**,** e19366.

DE LA CONCHA, E. G., CAVANILLAS, M. L., CENIT, M. C., URCELAY, E., ARROYO, R., FERNANDEZ, O., ALVAREZ-CERMENO, J. C., LEYVA, L., VILLAR, L. M. & NUNEZ, C. 2012. DRB1*03:01 haplotypes: differential contribution to multiple sclerosis risk and specific association with the presence of intrathecal IgM bands. *PLoS One,* 7**,** e31018.

DE VOS, J., HOSE, D., REME, T., TARTE, K., MOREAUX, J., MAHTOUK, K., JOURDAN, M., GOLDSCHMIDT, H., ROSSI, J. F., CREMER, F. W. & KLEIN, B. 2006. Microarray-based understanding of normal and malignant plasma cells. *Immunol Rev,* 210**,** 86-104.

DEAPEN, D., ESCALANTE, A., WEINRIB, L., HORWITZ, D., BACHMAN, B., ROY-BURMAN, P., WALKER, A. & MACK, T. M. 1992. A revised estimate of twin concordance in systemic lupus erythematosus. *Arthritis Rheum,* 35**,** 311-8.

DEGLI-ESPOSTI, M. A., ANDREAS, A., CHRISTIANSEN, F. T., SCHALKE, B., ALBERT, E. & DAWKINS, R. L. 1992. An approach to the localization of the susceptibility genes for generalized myasthenia gravis by mapping recombinant ancestral haplotypes. *Immunogenetics,* 35**,** 355-64.

DELGADO-VEGA, A. M., ALARCON-RIQUELME, M. E. & KOZYREV, S. V. 2010. Genetic associations in type I interferon related pathways with autoimmunity. *Arthritis Res Ther,* 12 Suppl 1**,** S2.

DENG, Y. & TSAO, B. P. 2010. Genetic susceptibility to systemic lupus erythematosus in the genomic era. *Nat Rev Rheumatol,* 6**,** 683-92.

DESCOTES, J. 2004. *Principles and Methods of Immunotoxicology*, Elsevier Science.

DILLON, S., AGGARWAL, R., HARDING, J. W., LI, L. J., WEISSMAN, M. H., LI, S., CAVETT, J. W., SEVIER, S. T., OJWANG, J. W., D'SOUZA, A., HARLEY, J. B. & SCOFIELD, R. H. 2011. Klinefelter's syndrome (47,XXY) among men with systemic lupus erythematosus. *Acta Paediatr,* 100**,** 819-23.

DING, B., PADYUKOV, L., LUNDSTROM, E., SEIELSTAD, M., PLENGE, R. M., OKSENBERG, J. R., GREGERSEN, P. K., ALFREDSSON, L. & KLARESKOG, L. 2009. Different patterns of associations with anti-citrullinated protein antibody-positive and anti-citrullinated protein antibody-negative rheumatoid arthritis in the extended major histocompatibility complex region. *Arthritis Rheum,* 60**,** 30-8.

DING, C., MA, Y., CHEN, X., LIU, M., CAI, Y., HU, X., XIANG, D., NATH, S., ZHANG, H. G., YE, H., POWELL, D. & YAN, J. 2013. Integrin CD11b negatively regulates BCR signalling to maintain autoreactive B cell tolerance. *Nat Commun,* 4**,** 2813.

ERSAHIN, T., CARKACIOGLU, L., CAN, T., KONU, O., ATALAY, V. & CETIN-ATALAY, R. 2014. Identification of novel reference genes based on MeSH categories. *PLoS One,* 9**,** e93341.

EVSYUKOVA, I., SOMARELLI, J. A., GREGORY, S. G. & GARCIA-BLANCO, M. A. 2010. Alternative splicing in multiple sclerosis and other autoimmune diseases. *RNA Biol,* 7**,** 462-73.

FAIRFAX, B. P., HUMBURG, P., MAKINO, S., NARANBHAI, V., WONG, D., LAU, E., JOSTINS, L., PLANT, K., ANDREWS, R., MCGEE, C. & KNIGHT, J. C. 2014. Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression. *Science,* 343**,** 1246949.

FAIRFAX, B. P. & KNIGHT, J. C. 2014. Genetics of gene expression in immunity to infection. *Curr Opin Immunol,* 30**,** 63-71.

FAIRFAX, B. P., MAKINO, S., RADHAKRISHNAN, J., PLANT, K., LESLIE, S., DILTHEY, A., ELLIS, P., LANGFORD, C., VANNBERG, F. O. & KNIGHT, J. C. 2012. Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles. *Nat Genet,* 44**,** 502-10.

FEHRMANN, R. S., JANSEN, R. C., VELDINK, J. H., WESTRA, H. J., ARENDS, D., BONDER, M. J., FU, J., DEELEN, P., GROEN, H. J., SMOLONSKA, A., WEERSMA, R. K., HOFSTRA, R. M., BUURMAN, W. A., RENSEN, S., WOLFS, M. G., PLATTEEL, M., ZHERNAKOVA, A., ELBERS, C. C., FESTEN, E. M., TRYNKA, G., HOFKER, M. H., SARIS, C. G., OPHOFF, R. A., VAN DEN BERG, L. H., VAN HEEL, D. A., WIJMENGA, C., TE MEERMAN, G. J. & FRANKE, L. 2011. Trans-eQTLs reveal that independent genetic variants associated with a complex phenotype converge on intermediate genes, with a major role for the HLA. *PLoS Genet,* 7**,** e1002197.

FENG, D., STONE, R. C., ELORANTA, M. L., SANGSTER-GUITY, N., NORDMARK, G., SIGURDSSON, S., WANG, C., ALM, G., SYVANEN, A. C., RONNBLOM, L. & BARNES, B. J. 2010. Genetic variants and disease-associated factors contribute to enhanced interferon regulatory factor 5 expression in blood cells of patients with systemic lupus erythematosus. *Arthritis Rheum,* 62**,** 562-73.

FERNANDO, M. M., FREUDENBERG, J., LEE, A., MORRIS, D. L., BOTEVA, L., RHODES, B., GONZALEZ-ESCRIBANO, M. F., LOPEZ-NEVOT, M. A., NAVARRA, S. V., GREGERSEN, P. K., MARTIN, J. & VYSE, T. J. 2012. Transancestral mapping of the MHC region in systemic lupus erythematosus identifies new independent and interacting loci at MSH5, HLA-DPB1 and HLA-G. *Ann Rheum Dis,* 71**,** 777-84.

FERNANDO, M. M., STEVENS, C. R., SABETI, P. C., WALSH, E. C., MCWHINNIE, A. J., SHAH, A., GREEN, T., RIOUX, J. D. & VYSE, T. J. 2007. Identification of two independent risk factors for lupus within the MHC in United Kingdom families. *PLoS Genet,* 3**,** e192.

FERNANDO, M. M., STEVENS, C. R., WALSH, E. C., DE JAGER, P. L., GOYETTE, P., PLENGE, R. M., VYSE, T. J. & RIOUX, J. D. 2008. Defining the role of the MHC in autoimmunity: a review and pooled analysis. *PLoS Genet,* 4**,** e1000024.

FERREIRA, R. C., GUO, H., COULSON, R. M., SMYTH, D. J., PEKALSKI, M. L., BURREN, O. S., CUTLER, A. J., DOECKE, J. D., FLINT, S., MCKINNEY, E. F., LYONS, P. A., SMITH, K. G., ACHENBACH, P., BEYERLEIN, A., DUNGER, D. B., CLAYTON, D. G., WICKER, L. S., TODD, J. A., BONIFACIO, E., WALLACE, C. & ZIEGLER, A. G. 2014. A type I interferon transcriptional signature precedes autoimmunity in children genetically at risk for type 1 diabetes. *Diabetes,* 63**,** 2538-50.

FINK, K., LANG, K. S., MANJARREZ-ORDUNO, N., JUNT, T., SENN, B. M., HOLDENER, M., AKIRA, S., ZINKERNAGEL, R. M. & HENGARTNER, H. 2006. Early type I interferon-mediated signals on B cells specifically enhance antiviral humoral responses. *Eur J Immunol,* 36**,** 2094-105.

FLESHER, D. L., SUN, X., BEHRENS, T. W., GRAHAM, R. R. & CRISWELL, L. A. 2010. Recent advances in the genetics of systemic lupus erythematosus. *Expert Rev Clin Immunol,* 6**,** 461-79.

FLOTO, R. A., CLATWORTHY, M. R., HEILBRONN, K. R., ROSNER, D. R., MACARY, P. A., RANKIN, A., LEHNER, P. J., OUWEHAND, W. H., ALLEN, J. M., WATKINS, N. A. & SMITH, K. G. 2005. Loss of function of a lupus-associated FcgammaRIIb polymorphism through exclusion from lipid rafts. *Nat Med,* 11**,** 1056-8.

FORABOSCO, P., GORMAN, J. D., CLEVELAND, C., KELLY, J. A., FISHER, S. A., ORTMANN, W. A., JOHANSSON, C., JOHANNESON, B., MOSER, K. L., GAFFNEY, P. M., TSAO, B. P., CANTOR, R. M., ALARCON-RIQUELME, M. E., BEHRENS, T. W., HARLEY, J. B., LEWIS, C. M. & CRISWELL, L. A. 2006. Meta-analysis of genome-wide linkage studies of systemic lupus erythematosus. *Genes Immun,* 7**,** 609-14.

FU, Q., ZHAO, J., QIAN, X., WONG, J. L., KAUFMAN, K. M., YU, C. Y., HWEE SIEW, H., TAN TOCK SENG HOSPITAL LUPUS STUDY, G., MOK, M. Y., HARLEY, J. B., GUTHRIDGE, J. M., SONG, Y. W., CHO, S. K., BAE, S. C., GROSSMAN, J. M., HAHN,

B. H., ARNETT, F. C., SHEN, N. & TSAO, B. P. 2011. Association of a functional IRF7 variant with systemic lupus erythematosus. *Arthritis Rheum,* 63**,** 749-54.

FURNROHR, B. G., WACH, S., KELLY, J. A., HASLBECK, M., WEBER, C. K., STACH, C. M., HUEBER, A. J., GRAEF, D., SPRIEWALD, B. M., MANGER, K., HERRMANN, M., KAUFMAN, K. M., FRANK, S. G., GOODMON, E., JAMES, J. A., SCHETT, G., WINKLER, T. H., HARLEY, J. B. & VOLL, R. E. 2010. Polymorphisms in the Hsp70 gene locus are genetically associated with systemic lupus erythematosus. *Ann Rheum Dis,* 69**,** 1983-9.

FURUKAWA, H., KAWASAKI, A., OKA, S., ITO, I., SHIMADA, K., SUGII, S., HASHIMOTO, A., KOMIYA, A., FUKUI, N., KONDO, Y., ITO, S., HAYASHI, T., MATSUMOTO, I., KUSAOI, M., AMANO, H., NAGAI, T., HIROHATA, S., SETOGUCHI, K., KONO, H., OKAMOTO, A., CHIBA, N., SUEMATSU, E., KATAYAMA, M., MIGITA, K., SUDA, A., OHNO, S., HASHIMOTO, H., TAKASAKI, Y., SUMIDA, T., NAGAOKA, S., TSUCHIYA, N. & TOHMA, S. 2014. Human leukocyte antigens and systemic lupus erythematosus: a protective role for the HLA-DR6 alleles DRB1*13:02 and *14:03. *PLoS One,* 9**,** e87792.

FYE, J. M., OREBAUGH, C. D., COFFIN, S. R., HOLLIS, T. & PERRINO, F. W. 2011. Dominant mutation of the TREX1 exonuclease gene in lupus and Aicardi-Goutieres syndrome. *J Biol Chem,* 286**,** 32373-82.

GAFFNEY, P. M., KEARNS, G. M., SHARK, K. B., ORTMANN, W. A., SELBY, S. A., MALMGREN, M. L., ROHLF, K. E., OCKENDEN, T. C., MESSNER, R. P., KING, R. A., RICH, S. S. & BEHRENS, T. W. 1998. A genome-wide search for susceptibility genes in human systemic lupus erythematosus sib-pair families. *Proc Natl Acad Sci U S A,* 95**,** 14875-9.

GAFFNEY, P. M., ORTMANN, W. A., SELBY, S. A., SHARK, K. B., OCKENDEN, T. C., ROHLF, K. E., WALGRAVE, N. L., BOYUM, W. P., MALMGREN, M. L., MILLER, M. E., KEARNS, G. M., MESSNER, R. P., KING, R. A., RICH, S. S. & BEHRENS, T. W. 2000. Genome screening in human systemic lupus erythematosus: results from a second Minnesota cohort and combined analyses of 187 sib-pair families. *Am J Hum Genet,* 66**,** 547-56.

GALEAZZI, M., SEBASTIANI, G. D., MOROZZI, G., CARCASSI, C., FERRARA, G. B., SCORZA, R., CERVERA, R., DE RAMON GARRIDO, E., FERNANDEZ-NEBRO, A., HOUSSIAU, F., JEDRYKA-GORAL, A., PASSIU, G., PAPASTERIADES, C., PIETTE, J. C., SMOLEN, J., PORCIELLO, G., MARCOLONGO, R. & EUROPEAN CONCERTEE ACTION ON THE IMMUNOGENETICS OF, S. L. E. 2002. HLA class II DNA typing in a large series of European patients with systemic lupus erythematosus: correlations with clinical and autoantibody subsets. *Medicine (Baltimore),* 81**,** 169-78.

GANTNER, F., HERMANN, P., NAKASHIMA, K., MATSUKAWA, S., SAKAI, K. & BACON, K. B. 2003. CD40-dependent and -independent activation of human tonsil B cells by CpG oligodeoxynucleotides. *Eur J Immunol,* 33**,** 1576-85.

GATEVA, V., SANDLING, J. K., HOM, G., TAYLOR, K. E., CHUNG, S. A., SUN, X., ORTMANN, W., KOSOY, R., FERREIRA, R. C., NORDMARK, G., GUNNARSSON, I., SVENUNGSSON, E., PADYUKOV, L., STURFELT, G., JONSEN, A., BENGTSSON, A. A., RANTAPAA-DAHLQVIST, S., BAECHLER, E. C., BROWN, E. E., ALARCON, G. S., EDBERG, J. C., RAMSEY-GOLDMAN, R., MCGWIN, G., JR., REVEILLE, J. D., VILA, L. M., KIMBERLY, R. P., MANZI, S., PETRI, M. A., LEE, A., GREGERSEN, P. K., SELDIN, M. F., RONNBLOM, L., CRISWELL, L. A., SYVANEN, A. C., BEHRENS, T. W. & GRAHAM, R. R. 2009. A large-scale replication study identifies TNIP1, PRDM1, JAZF1, UHRF1BP1 and IL10 as risk loci for systemic lupus erythematosus. *Nat Genet,* 41**,** 1228-33.

GIORDANI, L., SANCHEZ, M., LIBRI, I., QUARANTA, M. G., MATTIOLI, B. & VIORA, M. 2009. IFN-alpha amplifies human naive B cell TLR-9-mediated activation and Ig production. *J Leukoc Biol,* 86**,** 261-71.

GLUCKMAN, P. D. & HANSON, M. A. 2004. Living with the past: evolution, development, and patterns of disease. *Science,* 305**,** 1733-6.

GOLDBERG, M. A., ARNETT, F. C., BIAS, W. B. & SHULMAN, L. E. 1976. Histocompatibility antigens in systemic lupus erythematosus. *Arthritis Rheum,* 19**,** 129-32.

GOTTENBERG, J. E., BUSSON, M., LOISEAU, P., COHEN-SOLAL, J., LEPAGE, V., CHARRON, D., SIBILIA, J. & MARIETTE, X. 2003. In primary Sjogren's syndrome, HLA class II is associated exclusively with autoantibody production and spreading of the autoimmune response. *Arthritis Rheum,* 48**,** 2240-5.

GOURLEY, M. & MILLER, F. W. 2007. Mechanisms of disease: Environmental factors in the pathogenesis of rheumatic disease. *Nat Clin Pract Rheumatol,* 3**,** 172-80.

GRAHAM, R. R., COTSAPAS, C., DAVIES, L., HACKETT, R., LESSARD, C. J., LEON, J. M., BURTT, N. P., GUIDUCCI, C., PARKIN, M., GATES, C., PLENGE, R. M., BEHRENS, T. W., WITHER, J. E., RIOUX, J. D., FORTIN, P. R., GRAHAM, D. C., WONG, A. K., VYSE, T. J., DALY, M. J., ALTSHULER, D., MOSER, K. L. & GAFFNEY, P. M. 2008. Genetic variants near TNFAIP3 on 6q23 are associated with systemic lupus erythematosus. *Nat Genet,* 40**,** 1059-61.

GRAHAM, R. R., ORTMANN, W. A., LANGEFELD, C. D., JAWAHEER, D., SELBY, S. A., RODINE, P. R., BAECHLER, E. C., ROHLF, K. E., SHARK, K. B., ESPE, K. J., GREEN, L. E., NAIR, R. P., STUART, P. E., ELDER, J. T., KING, R. A., MOSER, K. L., GAFFNEY, P. M., BUGAWAN, T. L., ERLICH, H. A., RICH, S. S., GREGERSEN, P. K. & BEHRENS, T. W. 2002. Visualizing human leukocyte antigen class II risk haplotypes in human systemic lupus erythematosus. *Am J Hum Genet,* 71**,** 543-53.

GREENBERG, S. A., HIGGS, B. W., MOREHOUSE, C., WALSH, R. J., KONG, S. W., BROHAWN, P., ZHU, W., AMATO, A., SALAJEGHEH, M., WHITE, B., KIENER, P. A., JALLAL, B. & YAO, Y. 2012. Relationship between disease activity and type 1 interferon-and other cytokine-inducible gene expression in blood in dermatomyositis and polymyositis. *Genes and Immunity,* 13**,** 207-213.

GREGERSEN, P. K., DIAMOND, B. & PLENGE, R. M. 2012. GWAS implicates a role for quantitative immune traits and threshold effects in risk for human autoimmune disorders. *Curr Opin Immunol,* 24**,** 538-43.

GRIFFIN, D. O., QUACH, T., BATLIWALLA, F., ANDREOPOULOS, D., HOLODICK, N. E. & ROTHSTEIN, T. L. 2012. Human CD11b+ B1 cells are not monocytes: A reply to "Gene profiling of CD11b+ and CD11b- B1 cell subsets reveals potential cell sorting artifacts". *J Exp Med,* 209**,** 434-6.

GRIGORYEV, Y. A., KURIAN, S. M., NAKORCHEVSKIY, A. A., BURKE, J. P., CAMPBELL, D., HEAD, S. R., DENG, J., KANTOR, A. B., YATES, J. R., 3RD & SALOMON, D. R. 2009. Genome-wide analysis of immune activation in human T and B cells reveals distinct classes of alternatively spliced genes. *PLoS One,* 4**,** e7906.

GRUBIC, Z., ZUNEC, R., PEROS-GOLUBICIC, T., TEKAVEC-TRKANJEC, J., MARTINEZ, N., ALILOVIC, M., SMOJVER-JEZEK, S. & KERHIN-BRKLJACIC, V. 2007. HLA class I and class II frequencies in patients with sarcoidosis from Croatia: role of HLA-B8, -DRB1*0301, and -DQB1*0201 haplotype in clinical variations of the disease. *Tissue Antigens,* 70**,** 301-6.

GRUMET, F. C., COUKELL, A., BODMER, J. G., BODMER, W. F. & MCDEVITT, H. O. 1971. Histocompatibility (HL-A) antigens associated with systemic lupus erythematosus. A possible genetic predisposition to disease. *N Engl J Med,* 285**,** 193-6.

GRUNDBERG, E., SMALL, K. S., HEDMAN, A. K., NICA, A. C., BUIL, A., KEILDSON, S., BELL, J. T., YANG, T. P., MEDURI, E., BARRETT, A., NISBETT, J., SEKOWSKA, M., WILK, A., SHIN, S. Y., GLASS, D., TRAVERS, M., MIN, J. L., RING, S., HO, K., THORLEIFSSON, G., KONG, A., THORSTEINDOTTIR, U., AINALI, C., DIMAS, A. S., HASSANALI, N., INGLE, C., KNOWLES, D., KRESTYANINOVA, M., LOWE, C. E., DI MEGLIO, P., MONTGOMERY, S. B., PARTS, L., POTTER, S., SURDULESCU, G., TSAPROUNI, L., TSOKA, S., BATAILLE, V., DURBIN, R., NESTLE, F. O., O'RAHILLY, S., SORANZO, N., LINDGREN, C. M., ZONDERVAN, K. T., AHMADI, K. R., SCHADT, E. E., STEFANSSON, K., SMITH, G. D., MCCARTHY, M. I., DELOUKAS, P., DERMITZAKIS, E. T., SPECTOR, T. D. & MULTIPLE TISSUE HUMAN EXPRESSION RESOURCE, C. 2012. Mapping cis- and trans-regulatory effects across multiple tissues in twins. *Nat Genet,* 44**,** 1084-9.

GTEX CONSORTIUM 2015. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science,* 348**,** 648-60.

GUJER, C., SANDGREN, K. J., DOUAGI, I., ADAMS, W. C., SUNDLING, C., SMED-SORENSEN, A., SEDER, R. A., KARLSSON HEDESTAM, G. B. & LORE, K. 2011. IFN-alpha produced by human plasmacytoid dendritic cells enhances T cell-dependent naive B cell differentiation. *J Leukoc Biol,* 89**,** 811-21.

GUNTHER, C., KIND, B., REIJNS, M. A., BERNDT, N., MARTINEZ-BUENO, M., WOLF, C., TUNGLER, V., CHARA, O., LEE, Y. A., HUBNER, N., BICKNELL, L., BLUM, S., KRUG, C., SCHMIDT, F., KRETSCHMER, S., KOSS, S., ASTELL, K. R., RAMANTANI, G., BAUERFEIND, A., MORRIS, D. L., CUNNINGHAME GRAHAM, D. S., BUBECK, D., LEITCH, A., RALSTON, S. H., BLACKBURN, E. A., GAHR, M., WITTE, T., VYSE, T. J., MELCHERS, I., MANGOLD, E., NOTHEN, M. M., ARINGER, M., KUHN, A., LUTHKE, K., UNGER, L., BLEY, A., LORENZI, A., ISAACS, J. D., ALEXOPOULOU, D., CONRAD, K., DAHL, A., ROERS, A., ALARCON-RIQUELME, M. E., JACKSON, A. P.

& LEE-KIRSCH, M. A. 2015. Defective removal of ribonucleotides from DNA promotes systemic autoimmunity. *J Clin Invest,* 125**,** 413-24.
GUPTA, S., AGRAWAL, S. & GOLLAPUDI, S. 2013. Increased activation and cytokine secretion in B cells stimulated with leptin in aged humans. *Immun Ageing,* 10**,** 3.
HAN, J. W., ZHENG, H. F., CUI, Y., SUN, L. D., YE, D. Q., HU, Z., XU, J. H., CAI, Z. M., HUANG, W., ZHAO, G. P., XIE, H. F., FANG, H., LU, Q. J., XU, J. H., LI, X. P., PAN, Y. F., DENG, D. Q., ZENG, F. Q., YE, Z. Z., ZHANG, X. Y., WANG, Q. W., HAO, F., MA, L., ZUO, X. B., ZHOU, F. S., DU, W. H., CHENG, Y. L., YANG, J. Q., SHEN, S. K., LI, J., SHENG, Y. J., ZUO, X. X., ZHU, W. F., GAO, F., ZHANG, P. L., GUO, Q., LI, B., GAO, M., XIAO, F. L., QUAN, C., ZHANG, C., ZHANG, Z., ZHU, K. J., LI, Y., HU, D. Y., LU, W. S., HUANG, J. L., LIU, S. X., LI, H., REN, Y. Q., WANG, Z. X., YANG, C. J., WANG, P. G., ZHOU, W. M., LV, Y. M., ZHANG, A. P., ZHANG, S. Q., LIN, D., LI, Y., LOW, H. Q., SHEN, M., ZHAI, Z. F., WANG, Y., ZHANG, F. Y., YANG, S., LIU, J. J. & ZHANG, X. J. 2009. Genome-wide association study in a Chinese Han population identifies nine new susceptibility loci for systemic lupus erythematosus. *Nat Genet,* 41**,** 1234-7.
HARLY, C., GUILLAUME, Y., NEDELLEC, S., PEIGNE, C. M., MONKKONEN, H., MONKKONEN, J., LI, J., KUBALL, J., ADAMS, E. J., NETZER, S., DECHANET-MERVILLE, J., LEGER, A., HERRMANN, T., BREATHNACH, R., OLIVE, D., BONNEVILLE, M. & SCOTET, E. 2012. Key implication of CD277/butyrophilin-3 (BTN3A) in cellular stress sensing by a major human gammadelta T-cell subset. *Blood,* 120**,** 2269-79.
HELMIG, S., ALIAHMADI, N., STEPHAN, P., DOHREL, J. & SCHNEIDER, J. 2011. TNF-alpha -308 genotypes are associated with TNF-alpha and TGF-beta(1) mRNA expression in blood leucocytes of humans. *Cytokine,* 53**,** 306-10.
HERVAS-STUBBS, S., PEREZ-GRACIA, J. L., ROUZAUT, A., SANMAMED, M. F., LE BON, A. & MELERO, I. 2011. Direct effects of type I interferons on cells of the immune system. *Clin Cancer Res,* 17**,** 2619-27.
HINDORFF, L. A., SETHUPATHY, P., JUNKINS, H. A., RAMOS, E. M., MEHTA, J. P., COLLINS, F. S. & MANOLIO, T. A. 2009. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A,* 106**,** 9362-7.
HOCHBERG, M. C. 1997. Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus. *Arthritis Rheum,* 40**,** 1725.
HOFFMAN, B. & LIEBERMANN, D. A. 2008. Apoptotic signaling by c-MYC. *Oncogene,* 27**,** 6462-72.
HOFFMANN, H. H., SCHNEIDER, W. M. & RICE, C. M. 2015. Interferons and viruses: an evolutionary arms race of molecular interactions. *Trends Immunol,* 36**,** 124-38.
HOLMES, D. A., SUTO, E., LEE, W. P., OU, Q., GONG, Q., SMITH, H. R., CAPLAZI, P. & CHAN, A. C. 2015. Autoimmunity-associated protein tyrosine phosphatase PEP negatively regulates IFN-alpha receptor signaling. *J Exp Med,* 212**,** 1081-93.
HOM, G., GRAHAM, R. R., MODREK, B., TAYLOR, K. E., ORTMANN, W., GARNIER, S., LEE, A. T., CHUNG, S. A., FERREIRA, R. C., PANT, P. V., BALLINGER, D. G., KOSOY, R., DEMIRCI, F. Y., KAMBOH, M. I., KAO, A. H., TIAN, C., GUNNARSSON, I., BENGTSSON, A. A., RANTAPAA-DAHLQVIST, S., PETRI, M., MANZI, S., SELDIN, M. F., RONNBLOM, L., SYVANEN, A. C., CRISWELL, L. A., GREGERSEN, P. K. & BEHRENS, T. W. 2008. Association of systemic lupus erythematosus with C8orf13-BLK and ITGAM-ITGAX. *N Engl J Med,* 358**,** 900-9.
HOOKS, J. J., MOUTSOPOULOS, H. M., GEIS, S. A., STAHL, N. I., DECKER, J. L. & NOTKINS, A. L. 1979. Immune interferon in the circulation of patients with autoimmune disease. *N Engl J Med,* 301**,** 5-8.
HORNBERG, M., AROLT, V., WILKE, I., KRUSE, A. & KIRCHNER, H. 1995. Production of interferons and lymphokines in leukocyte cultures of patients with schizophrenia. *Schizophr Res,* 15**,** 237-42.
HORTON, R., GIBSON, R., COGGILL, P., MIRETTI, M., ALLCOCK, R. J., ALMEIDA, J., FORBES, S., GILBERT, J. G., HALLS, K., HARROW, J. L., HART, E., HOWE, K., JACKSON, D. K., PALMER, S., ROBERTS, A. N., SIMS, S., STEWART, C. A., TRAHERNE, J. A., TREVANION, S., WILMING, L., ROGERS, J., DE JONG, P. J., ELLIOTT, J. F., SAWCER, S., TODD, J. A., TROWSDALE, J. & BECK, S. 2008. Variation analysis and gene annotation of eight MHC haplotypes: the MHC Haplotype Project. *Immunogenetics,* 60**,** 1-18.
HORTON, R., WILMING, L., RAND, V., LOVERING, R. C., BRUFORD, E. A., KHODIYAR, V. K., LUSH, M. J., POVEY, S., TALBOT, C. C., JR., WRIGHT, M. W., WAIN, H. M.,

TROWSDALE, J., ZIEGLER, A. & BECK, S. 2004. Gene map of the extended human MHC. *Nat Rev Genet,* 5**,** 889-99.

HRUZ, T., WYSS, M., DOCQUIER, M., PFAFFL, M. W., MASANETZ, S., BORGHI, L., VERBRUGGHE, P., KALAYDJIEVA, L., BLEULER, S., LAULE, O., DESCOMBES, P., GRUISSEM, W. & ZIMMERMANN, P. 2011. RefGenes: identification of reliable and condition specific reference genes for RT-qPCR data normalization. *BMC Genomics,* 12**,** 156.

HU, S. J., WEN, L. L., HU, X., YIN, X. Y., CUI, Y., YANG, S. & ZHANG, X. J. 2013. IKZF1: a critical role in the pathogenesis of systemic lupus erythematosus? *Mod Rheumatol,* 23**,** 205-9.

INGLOT, A. D., LESZEK, J., PIASECKI, E. & SYPULA, A. 1994. Interferon responses in schizophrenia and major depressive disorders. *Biol Psychiatry,* 35**,** 464-73.

INTERNATIONAL CONSORTIUM FOR SYSTEMIC LUPUS ERYTHEMATOSUS, G., HARLEY, J. B., ALARCON-RIQUELME, M. E., CRISWELL, L. A., JACOB, C. O., KIMBERLY, R. P., MOSER, K. L., TSAO, B. P., VYSE, T. J., LANGEFELD, C. D., NATH, S. K., GUTHRIDGE, J. M., COBB, B. L., MIREL, D. B., MARION, M. C., WILLIAMS, A. H., DIVERS, J., WANG, W., FRANK, S. G., NAMJOU, B., GABRIEL, S. B., LEE, A. T., GREGERSEN, P. K., BEHRENS, T. W., TAYLOR, K. E., FERNANDO, M., ZIDOVETZKI, R., GAFFNEY, P. M., EDBERG, J. C., RIOUX, J. D., OJWANG, J. O., JAMES, J. A., MERRILL, J. T., GILKESON, G. S., SELDIN, M. F., YIN, H., BAECHLER, E. C., LI, Q. Z., WAKELAND, E. K., BRUNER, G. R., KAUFMAN, K. M. & KELLY, J. A. 2008. Genome-wide association scan in women with systemic lupus erythematosus identifies susceptibility variants in ITGAM, PXK, KIAA1542 and other loci. *Nat Genet,* 40**,** 204-10.

INTERNATIONAL MHC AND AUTOIMMUNITY GENETICS NETWORK, RIOUX, J. D., GOYETTE, P., VYSE, T. J., HAMMARSTROM, L., FERNANDO, M. M., GREEN, T., DE JAGER, P. L., FOISY, S., WANG, J., DE BAKKER, P. I., LESLIE, S., MCVEAN, G., PADYUKOV, L., ALFREDSSON, L., ANNESE, V., HAFLER, D. A., PAN-HAMMARSTROM, Q., MATELL, R., SAWCER, S. J., COMPSTON, A. D., CREE, B. A., MIREL, D. B., DALY, M. J., BEHRENS, T. W., KLARESKOG, L., GREGERSEN, P. K., OKSENBERG, J. R. & HAUSER, S. L. 2009. Mapping of multiple susceptibility variants within the MHC region for 7 immune-mediated diseases. *Proc Natl Acad Sci U S A,* 106**,** 18680-5.

JACOB, C. O., EISENSTEIN, M., DINAUER, M. C., MING, W., LIU, Q., JOHN, S., QUISMORIO, F. P., JR., REIFF, A., MYONES, B. L., KAUFMAN, K. M., MCCURDY, D., HARLEY, J. B., SILVERMAN, E., KIMBERLY, R. P., VYSE, T. J., GAFFNEY, P. M., MOSER, K. L., KLEIN-GITELMAN, M., WAGNER-WEINER, L., LANGEFELD, C. D., ARMSTRONG, D. L. & ZIDOVETZKI, R. 2012. Lupus-associated causal mutation in neutrophil cytosolic factor 2 (NCF2) brings unique insights to the structure and function of NADPH oxidase. *Proc Natl Acad Sci U S A,* 109**,** E59-67.

JARJOUR, W., REED, A. M., GAUTHIER, J., HUNT, S., 3RD & WINFIELD, J. B. 1996. The 8.5-kb PstI allele of the stress protein gene, Hsp70-2: an independent risk factor for systemic lupus erythematosus in African Americans? *Hum Immunol,* 45**,** 59-63.

JEFFRIES, M. A. & SAWALHA, A. H. 2011. Epigenetics in systemic lupus erythematosus: leading the way for specific therapeutic agents. *Int J Clin Rheumtol,* 6**,** 423-439.

JORES, R. D., FRAU, F., CUCCA, F., GRAZIA CLEMENTE, M., ORRU, S., RAIS, M., DE VIRGILIIS, S. & CONGIA, M. 2007. HLA-DQB1*0201 homozygosis predisposes to severe intestinal damage in celiac disease. *Scand J Gastroenterol,* 42**,** 48-53.

KALLENBERG, C. G., VAN DER VOORT-BEELEN, J. M., D'AMARO, J. & THE, T. H. 1981. Increased frequency of B8/DR3 in scleroderma and association of the haplotype with impaired cellular immune response. *Clin Exp Immunol,* 43**,** 478-85.

KAPITANY, A., TARR, T., GYETVAI, A., SZODORAY, P., TUMPEK, J., POOR, G., SZEGEDI, G., SIPKA, S. & KISS, E. 2009. Human leukocyte antigen-DRB1 and -DQB1 genotyping in lupus patients with and without antiphospholipid syndrome. *Ann N Y Acad Sci,* 1173**,** 545-51.

KARIUKI, S. N., KIROU, K. A., MACDERMOTT, E. J., BARILLAS-ARIAS, L., CROW, M. K. & NIEWOLD, T. B. 2009. Cutting edge: autoimmune disease risk variant of STAT4 confers increased sensitivity to IFN-alpha in lupus patients in vivo. *J Immunol,* 182**,** 34-8.

KATILA, H., CANTELL, K., HIRVONEN, S. & RIMON, R. 1989. Production of interferon-alpha and gamma by leukocytes from patients with schizophrenia. *Schizophr Res,* 2**,** 361-5.

KAWAI, K., TSUNO, N. H., MATSUHASHI, M., KITAYAMA, J., OSADA, T., YAMADA, J., TSUCHIYA, T., YONEYAMA, S., WATANABE, T., TAKAHASHI, K. & NAGAWA, H.

2005. CD11b-mediated migratory property of peripheral blood B cells. *J Allergy Clin Immunol,* 116**,** 192-7.

KELLEY, J. M., EDBERG, J. C. & KIMBERLY, R. P. 2010. Pathways: Strategies for susceptibility genes in SLE. *Autoimmun Rev,* 9**,** 473-6.

KIM, I., KIM, Y. J., KIM, K., KANG, C., CHOI, C. B., SUNG, Y. K., LEE, H. S. & BAE, S. C. 2009. Genetic studies of systemic lupus erythematosus in Asia: where are we now? *Genes Immun,* 10**,** 421-32.

KIM, K., BANG, S. Y., LEE, H. S., OKADA, Y., HAN, B., SAW, W. Y., TEO, Y. Y. & BAE, S. C. 2014. The HLA-DRbeta1 amino acid positions 11-13-26 explain the majority of SLE-MHC associations. *Nat Commun,* 5**,** 5902.

KIROU, K. A., LEE, C., GEORGE, S., LOUCA, K., PETERSON, M. G. & CROW, M. K. 2005. Activation of the interferon-alpha pathway identifies a subgroup of systemic lupus erythematosus patients with distinct serologic features and active disease. *Arthritis Rheum,* 52**,** 1491-503.

KRETSCHMER, S., WOLF, C., KONIG, N., STAROSKE, W., GUCK, J., HAUSLER, M., LUKSCH, H., NGUYEN, L. A., KIM, B., ALEXOPOULOU, D., DAHL, A., RAPP, A., CARDOSO, M. C., SHEVCHENKO, A. & LEE-KIRSCH, M. A. 2015. SAMHD1 prevents autoimmunity by maintaining genome stability. *Ann Rheum Dis,* 74**,** e17.

KUBLER, K., ARNDT, P. F., WARDELMANN, E., KREBS, D., KUHN, W. & VAN DER VEN, K. 2006. HLA-class II haplotype associations with ovarian cancer. *Int J Cancer,* 119**,** 2980-5.

KYOGOKU, C., SMILJANOVIC, B., GRUN, J. R., BIESEN, R., SCHULTE-WREDE, U., HAUPL, T., HIEPE, F., ALEXANDER, T., RADBRUCH, A. & GRUTZKAU, A. 2013. Cell-specific type I IFN signatures in autoimmunity and viral infection: what makes the difference? *PLoS One,* 8**,** e83776.

LAHITA, R. G. 2004. *Systemic Lupus Erythematosus*, Elsevier Science.

LALONDE, E., HA, K. C., WANG, Z., BEMMO, A., KLEINMAN, C. L., KWAN, T., PASTINEN, T. & MAJEWSKI, J. 2011. RNA sequencing reveals the role of splicing polymorphisms in regulating human gene expression. *Genome Res,* 21**,** 545-54.

LAPPALAINEN, T., SAMMETH, M., FRIEDLANDER, M. R., T HOEN, P. A., MONLONG, J., RIVAS, M. A., GONZALEZ-PORTA, M., KURBATOVA, N., GRIEBEL, T., FERREIRA, P. G., BARANN, M., WIELAND, T., GREGER, L., VAN ITERSON, M., ALMLOF, J., RIBECA, P., PULYAKHINA, I., ESSER, D., GIGER, T., TIKHONOV, A., SULTAN, M., BERTIER, G., MACARTHUR, D. G., LEK, M., LIZANO, E., BUERMANS, H. P., PADIOLEAU, I., SCHWARZMAYR, T., KARLBERG, O., ONGEN, H., KILPINEN, H., BELTRAN, S., GUT, M., KAHLEM, K., AMSTISLAVSKIY, V., STEGLE, O., PIRINEN, M., MONTGOMERY, S. B., DONNELLY, P., MCCARTHY, M. I., FLICEK, P., STROM, T. M., GEUVADIS, C., LEHRACH, H., SCHREIBER, S., SUDBRAK, R., CARRACEDO, A., ANTONARAKIS, S. E., HASLER, R., SYVANEN, A. C., VAN OMMEN, G. J., BRAZMA, A., MEITINGER, T., ROSENSTIEL, P., GUIGO, R., GUT, I. G., ESTIVILL, X. & DERMITZAKIS, E. T. 2013. Transcriptome and genome sequencing uncovers functional variation in humans. *Nature,* 501**,** 506-11.

LARSEN, C. E., ALFORD, D. R., TRAUTWEIN, M. R., JALLOH, Y. K., TARNACKI, J. L., KUNNENKERI, S. K., FICI, D. A., YUNIS, E. J., AWDEH, Z. L. & ALPER, C. A. 2014. Dominant sequences of human major histocompatibility complex conserved extended haplotypes from HLA-DQA2 to DAXX. *PLoS Genet,* 10**,** e1004637.

LAWRENCE, J. S., MARTINS, C. L. & DRAKE, G. L. 1987. A family survey of lupus erythematosus. 1. Heritability. *J Rheumatol,* 14**,** 913-21.

LE, K., MITSOURAS, K., ROY, M., WANG, Q., XU, Q., NELSON, S. F. & LEE, C. 2004. Detecting tissue-specific regulation of alternative splicing as a qualitative change in microarray data. *Nucleic Acids Res,* 32**,** e180.

LEE-KIRSCH, M. A., GONG, M., CHOWDHURY, D., SENENKO, L., ENGEL, K., LEE, Y. A., DE SILVA, U., BAILEY, S. L., WITTE, T., VYSE, T. J., KERE, J., PFEIFFER, C., HARVEY, S., WONG, A., KOSKENMIES, S., HUMMEL, O., ROHDE, K., SCHMIDT, R. E., DOMINICZAK, A. F., GAHR, M., HOLLIS, T., PERRINO, F. W., LIEBERMAN, J. & HUBNER, N. 2007. Mutations in the gene encoding the 3'-5' DNA exonuclease TREX1 are associated with systemic lupus erythematosus. *Nat Genet,* 39**,** 1065-7.

LEE, M. N., YE, C., VILLANI, A. C., RAJ, T., LI, W., EISENHAURE, T. M., IMBOYWA, S. H., CHIPENDO, P. I., RAN, F. A., SLOWIKOWSKI, K., WARD, L. D., RADDASSI, K., MCCABE, C., LEE, M. H., FROHLICH, I. Y., HAFLER, D. A., KELLIS, M., RAYCHAUDHURI, S., ZHANG, F., STRANGER, B. E., BENOIST, C. O., DE JAGER, P. L., REGEV, A. & HACOHEN, N. 2014. Common genetic variants modulate pathogen-sensing responses in human dendritic cells. *Science,* 343**,** 1246980.

LEEK, J. T., JOHNSON, W. E., PARKER, H. S., JAFFE, A. E. & STOREY, J. D. 2012. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics,* 28**,** 882-3.

LI, Y., SUN, L. D., LU, W. S., HU, W. L., GAO, J. P., CHENG, Y. L., YU, Z. Y., YAO, S., HE, C. F., LIU, J. L., CUI, Y. & YANG, S. 2010. Expression analysis of ETS1 gene in peripheral blood mononuclear cells with systemic lupus erythematosus by real-time reverse transcription PCR. *Chin Med J (Engl),* 123**,** 2287-8.

LIE, B. A. & THORSBY, E. 2005. Several genes in the extended human MHC contribute to predisposition to autoimmune diseases. *Curr Opin Immunol,* 17**,** 526-31.

LIMA-JUNIOR JDA, C. & PRATT-RICCIO, L. R. 2016. Major Histocompatibility Complex and Malaria: Focus on Plasmodium vivax Infection. *Front Immunol,* 7**,** 13.

LINCOLN, M. R., RAMAGOPALAN, S. V., CHAO, M. J., HERRERA, B. M., DELUCA, G. C., ORTON, S. M., DYMENT, D. A., SADOVNICK, A. D. & EBERS, G. C. 2009. Epistasis among HLA-DRB1, HLA-DQA1, and HLA-DQB1 loci determines multiple sclerosis susceptibility. *Proc Natl Acad Sci U S A,* 106**,** 7542-7.

LIOU, L. Y., BLASIUS, A. L., WELCH, M. J., COLONNA, M., OLDSTONE, M. B. & ZUNIGA, E. I. 2008. In vivo conversion of BM plasmacytoid DC into CD11b+ conventional DC during virus infection. *Eur J Immunol,* 38**,** 3388-94.

LISNEVSKAIA, L., MURPHY, G. & ISENBERG, D. 2014. Systemic lupus erythematosus. *Lancet,* 384**,** 1878-88.

LIU, Z., BETHUNAICKAN, R., HUANG, W., LODHI, U., SOLANO, I., MADAIO, M. P. & DAVIDSON, A. 2011. Interferon-alpha accelerates murine systemic lupus erythematosus in a T cell-dependent manner. *Arthritis Rheum,* 63**,** 219-29.

LIU, Z. & DAVIDSON, A. 2012. Taming lupus-a new understanding of pathogenesis is leading to clinical advances. *Nat Med,* 18**,** 871-82.

LOCKSHIN, M. D., LEVINE, A. B. & ERKAN, D. 2015. Patients with overlap autoimmune disease differ from those with 'pure' disease. *Lupus Sci Med,* 2**,** e000084.

LOCKSTONE, H. E. 2011. Exon array data analysis using Affymetrix power tools and R statistical software. *Brief Bioinform,* 12**,** 634-44.

LU, L. Y., DING, W. Z., FICI, D., DEULOFEUT, R., CHENG, H. H., CHEU, C. C., SUNG, P. K., SCHUR, P. H. & FRASER, P. A. 1997. Molecular analysis of major histocompatibility complex allelic associations with systemic lupus erythematosus in Taiwan. *Arthritis Rheum,* 40**,** 1138-45.

LU, X., ZOLLER, E. E., WEIRAUCH, M. T., WU, Z., NAMJOU, B., WILLIAMS, A. H., ZIEGLER, J. T., COMEAU, M. E., MARION, M. C., GLENN, S. B., ADLER, A., SHEN, N., NATH, S. K., STEVENS, A. M., FREEDMAN, B. I., TSAO, B. P., JACOB, C. O., KAMEN, D. L., BROWN, E. E., GILKESON, G. S., ALARCON, G. S., REVEILLE, J. D., ANAYA, J. M., JAMES, J. A., SIVILS, K. L., CRISWELL, L. A., VILA, L. M., ALARCON-RIQUELME, M. E., PETRI, M., SCOFIELD, R. H., KIMBERLY, R. P., RAMSEY-GOLDMAN, R., JOO, Y. B., CHOI, J., BAE, S. C., BOACKLE, S. A., GRAHAM, D. C., VYSE, T. J., GUTHRIDGE, J. M., GAFFNEY, P. M., LANGEFELD, C. D., KELLY, J. A., GREIS, K. D., KAUFMAN, K. M., HARLEY, J. B. & KOTTYAN, L. C. 2015. Lupus Risk Variant Increases pSTAT1 Binding and Decreases ETS1 Expression. *Am J Hum Genet,* 96**,** 731-9.

LUND, F. E. & RANDALL, T. D. 2010. Effector and regulatory B cells: modulators of CD4+ T cell immunity. *Nat Rev Immunol,* 10**,** 236-47.

LUNDBERG, I. E. & HELMERS, S. B. 2010. The type I interferon system in idiopathic inflammatory myopathies. *Autoimmunity,* 43**,** 239-43.

LUO, X., YANG, W., YE, D. Q., CUI, H., ZHANG, Y., HIRANKARN, N., QIAN, X., TANG, Y., LAU, Y. L., DE VRIES, N., TAK, P. P., TSAO, B. P. & SHEN, N. 2011. A functional variant in microRNA-146a promoter modulates its expression and confers disease risk for systemic lupus erythematosus. *PLoS Genet,* 7**,** e1002128.

MACIEL, L. M., RODRIGUES, S. S., DIBBERN, R. S., NAVARRO, P. A. & DONADI, E. A. 2001. Association of the HLA-DRB1*0301 and HLA-DQA1*0501 alleles with Graves' disease in a population representing the gene contribution from several ethnic backgrounds. *Thyroid,* 11**,** 31-5.

MANDON-PEPIN, B., TOURAINE, P., KUTTENN, F., DERBOIS, C., ROUXEL, A., MATSUDA, F., NICOLAS, A., COTINOT, C. & FELLOUS, M. 2008. Genetic investigation of four meiotic genes in women with premature ovarian failure. *Eur J Endocrinol,* 158**,** 107-15.

MANKU, H., LANGEFELD, C. D., GUERRA, S. G., MALIK, T. H., ALARCON-RIQUELME, M., ANAYA, J. M., BAE, S. C., BOACKLE, S. A., BROWN, E. E., CRISWELL, L. A., FREEDMAN, B. I., GAFFNEY, P. M., GREGERSEN, P. A., GUTHRIDGE, J. M., HAN, S. H., HARLEY, J. B., JACOB, C. O., JAMES, J. A., KAMEN, D. L., KAUFMAN, K. M.,

KELLY, J. A., MARTIN, J., MERRILL, J. T., MOSER, K. L., NIEWOLD, T. B., PARK, S. Y., PONS-ESTEL, B. A., SAWALHA, A. H., SCOFIELD, R. H., SHEN, N., STEVENS, A. M., SUN, C., GILKESON, G. S., EDBERG, J. C., KIMBERLY, R. P., NATH, S. K., TSAO, B. P. & VYSE, T. J. 2013. Trans-ancestral studies fine map the SLE-susceptibility locus TNFSF4. *PLoS Genet,* 9**,** e1003554.

MARTIN, J. E., ASSASSI, S., DIAZ-GALLO, L. M., BROEN, J. C., SIMEON, C. P., CASTELLVI, I., VICENTE-RABANEDA, E., FONOLLOSA, V., ORTEGO-CENTENO, N., GONZALEZ-GAY, M. A., ESPINOSA, G., CARREIRA, P., SPANISH SCLERODERMA, G., CONSORTIUM, S., GROUP, U. S. S. G., BIOLUPUS, CAMPS, M., SABIO, J. M., D'ALFONSO, S., VONK, M. C., VOSKUYL, A. E., SCHUERWEGH, A. J., KREUTER, A., WITTE, T., RIEMEKASTEN, G., HUNZELMANN, N., AIRO, P., BERETTA, L., SCORZA, R., LUNARDI, C., VAN LAAR, J., CHEE, M. M., WORTHINGTON, J., HERRICK, A., DENTON, C., FONSECA, C., TAN, F. K., ARNETT, F., ZHOU, X., REVEILLE, J. D., GORLOVA, O., KOELEMAN, B. P., RADSTAKE, T. R., VYSE, T., MAYES, M. D., ALARCON-RIQUELME, M. E. & MARTIN, J. 2013. A systemic sclerosis and systemic lupus erythematosus pan-meta-GWAS reveals new shared susceptibility loci. *Hum Mol Genet,* 22**,** 4021-9.

MAVRAGANI, C. P., LA, D. T., STOHL, W. & CROW, M. K. 2010. Association of the response to tumor necrosis factor antagonists with plasma type I interferon activity and interferon-beta/alpha ratios in rheumatoid arthritis patients: a post hoc analysis of a predominantly Hispanic cohort. *Arthritis Rheum,* 62**,** 392-401.

MAVRAGANI, C. P., NIEWOLD, T. B., MOUTSOPOULOS, N. M., PILLEMER, S. R., WAHL, S. M. & CROW, M. K. 2007. Augmented interferon-alpha pathway activation in patients with Sjogren's syndrome treated with etanercept. *Arthritis Rheum,* 56**,** 3995-4004.

MERRILL, J. T., WALLACE, D. J., PETRI, M., KIROU, K. A., YAO, Y., WHITE, W. I., ROBBIE, G., LEVIN, R., BERNEY, S. M., CHINDALORE, V., OLSEN, N., RICHMAN, L., LE, C., JALLAL, B., WHITE, B. & LUPUS INTERFERON SKIN ACTIVITY STUDY, I. 2011. Safety profile and clinical activity of sifalimumab, a fully human anti-interferon alpha monoclonal antibody, in systemic lupus erythematosus: a phase I, multicentre, double-blind randomised study. *Ann Rheum Dis,* 70**,** 1905-13.

MESSAL, N., MAMESSIER, E., SYLVAIN, A., CELIS-GUTIERREZ, J., THIBULT, M. L., CHETAILLE, B., FIRAGUAY, G., PASTOR, S., GUILLAUME, Y., WANG, Q., HIRSCH, I., NUNES, J. A. & OLIVE, D. 2011. Differential role for CD277 as a co-regulator of the immune signal in T and NK cells. *Eur J Immunol,* 41**,** 3443-54.

MILLER, F. W., CHEN, W., O'HANLON, T. P., COOPER, R. G., VENCOVSKY, J., RIDER, L. G., DANKO, K., WEDDERBURN, L. R., LUNDBERG, I. E., PACHMAN, L. M., REED, A. M., YTTERBERG, S. R., PADYUKOV, L., SELVA-O'CALLAGHAN, A., RADSTAKE, T. R., ISENBERG, D. A., CHINOY, H., OLLIER, W. E., SCHEET, P., PENG, B., LEE, A., BYUN, J., LAMB, J. A., GREGERSEN, P. K., AMOS, C. I. & MYOSITIS GENETICS, C. 2015. Genome-wide association study identifies HLA 8.1 ancestral haplotype alleles as major genetic risk factors for myositis phenotypes. *Genes Immun,* 16**,** 470-80.

MINER, J. J. & DIAMOND, M. S. 2014. MDA5 and autoimmune disease. *Nat Genet,* 46**,** 418-9.

MOHAN, C. & PUTTERMAN, C. 2015. Genetics and pathogenesis of systemic lupus erythematosus and lupus nephritis. *Nat Rev Nephrol,* 11**,** 329-41.

MORRIS, D. L., FERNANDO, M. M., TAYLOR, K. E., CHUNG, S. A., NITITHAM, J., ALARCON-RIQUELME, M. E., BARCELLOS, L. F., BEHRENS, T. W., COTSAPAS, C., GAFFNEY, P. M., GRAHAM, R. R., PONS-ESTEL, B. A., GREGERSEN, P. K., HARLEY, J. B., HAUSER, S. L., HOM, G., LANGEFELD, C. D., NOBLE, J. A., RIOUX, J. D., SELDIN, M. F., SYSTEMIC LUPUS ERYTHEMATOSUS GENETICS, C., VYSE, T. J. & CRISWELL, L. A. 2014. MHC associations with clinical and autoantibody manifestations in European SLE. *Genes Immun,* 15**,** 210-7.

MORRIS, D. L., TAYLOR, K. E., FERNANDO, M. M., NITITHAM, J., ALARCON-RIQUELME, M. E., BARCELLOS, L. F., BEHRENS, T. W., COTSAPAS, C., GAFFNEY, P. M., GRAHAM, R. R., PONS-ESTEL, B. A., GREGERSEN, P. K., HARLEY, J. B., HAUSER, S. L., HOM, G., LANGEFELD, C. D., NOBLE, J. A., RIOUX, J. D., SELDIN, M. F., CRISWELL, L. A. & VYSE, T. J. 2012. Unraveling multiple MHC gene associations with systemic lupus erythematosus: model choice indicates a role for HLA alleles and non-HLA genes in Europeans. *Am J Hum Genet,* 91**,** 778-93.

MUELLER, M., BARROS, P., WITHERDEN, A. S., ROBERTS, A. L., ZHANG, Z., SCHASCHL, H., YU, C. Y., HURLES, M. E., SCHAFFNER, C., FLOTO, R. A., GAME, L., STEINBERG, K. M., WILSON, R. K., GRAVES, T. A., EICHLER, E. E., COOK, H. T., VYSE, T. J. & AITMAN, T. J. 2013. Genomic pathology of SLE-associated copy-number variation at the FCGR2C/FCGR3B/FCGR2B locus. *Am J Hum Genet,* 92**,** 28-40.

MUSTAFA, M. Z., SCHOFIELD, J., MILLS, P. R., PRIEST, M., FOX, R., DATTA, S., MORRIS, J., FORREST, E. H., GILLESPIE, R., STANLEY, A. J. & BARCLAY, S. T. 2014. The efficacy and safety of treating hepatitis C in patients with a diagnosis of schizophrenia. *J Viral Hepat,* 21**,** e48-51.

NAMJOU, B., KOTHARI, P. H., KELLY, J. A., GLENN, S. B., OJWANG, J. O., ADLER, A., ALARCON-RIQUELME, M. E., GALLANT, C. J., BOACKLE, S. A., CRISWELL, L. A., KIMBERLY, R. P., BROWN, E., EDBERG, J., STEVENS, A. M., JACOB, C. O., TSAO, B. P., GILKESON, G. S., KAMEN, D. L., MERRILL, J. T., PETRI, M., GOLDMAN, R. R., VILA, L. M., ANAYA, J. M., NIEWOLD, T. B., MARTIN, J., PONS-ESTEL, B. A., SABIO, J. M., CALLEJAS, J. L., VYSE, T. J., BAE, S. C., PERRINO, F. W., FREEDMAN, B. I., SCOFIELD, R. H., MOSER, K. L., GAFFNEY, P. M., JAMES, J. A., LANGEFELD, C. D., KAUFMAN, K. M., HARLEY, J. B. & ATKINSON, J. P. 2011. Evaluation of the TREX1 gene in a large multi-ancestral lupus cohort. *Genes Immun,* 12**,** 270-9.

NARANBHAI, V., FAIRFAX, B. P., MAKINO, S., HUMBURG, P., WONG, D., NG, E., HILL, A. V. & KNIGHT, J. C. 2015. Genomic modulators of gene expression in human neutrophils. *Nat Commun,* 6**,** 7545.

NIEDERWIESER, D., AUBOCK, J., TROPPMAIR, J., HEROLD, M., SCHULER, G., BOECK, G., LOTZ, J., FRITSCH, P. & HUBER, C. 1988. IFN-mediated induction of MHC antigen expression on human keratinocytes and its influence on in vitro alloimmune responses. *J Immunol,* 140**,** 2556-64.

NIEWOLD, T. B., HUA, J., LEHMAN, T. J., HARLEY, J. B. & CROW, M. K. 2007. High serum IFN-alpha activity is a heritable risk factor for systemic lupus erythematosus. *Genes Immun,* 8**,** 492-502.

NIEWOLD, T. B., KELLY, J. A., FLESCH, M. H., ESPINOZA, L. R., HARLEY, J. B. & CROW, M. K. 2008. Association of the IRF5 risk haplotype with high serum interferon-alpha activity in systemic lupus erythematosus patients. *Arthritis Rheum,* 58**,** 2481-7.

NIEWOLD, T. B. & SWEDLER, W. I. 2005. Systemic lupus erythematosus arising during interferon-alpha therapy for cryoglobulinemic vasculitis associated with hepatitis C. *Clin Rheumatol,* 24**,** 178-81.

NOBLE, J. A. & VALDES, A. M. 2011. Genetics of the HLA region in the prediction of type 1 diabetes. *Curr Diab Rep,* 11**,** 533-42.

O'HANLON, T. P., CARRICK, D. M., TARGOFF, I. N., ARNETT, F. C., REVEILLE, J. D., CARRINGTON, M., GAO, X., ODDIS, C. V., MOREL, P. A., MALLEY, J. D., MALLEY, K., SHAMIM, E. A., RIDER, L. G., CHANOCK, S. J., FOSTER, C. B., BUNCH, T., BLACKSHEAR, P. J., PLOTZ, P. H., LOVE, L. A. & MILLER, F. W. 2006. Immunogenetic risk and protective factors for the idiopathic inflammatory myopathies: distinct HLA-A, -B, -Cw, -DRB1, and -DQA1 allelic profiles distinguish European American patients with different myositis autoantibodies. *Medicine (Baltimore),* 85**,** 111-27.

O'NEILL, S. & CERVERA, R. 2010. Systemic lupus erythematosus. *Best Pract Res Clin Rheumatol,* 24**,** 841-55.

OLIVEIRA, L. C., PORTA, G., MARIN, M. L., BITTENCOURT, P. L., KALIL, J. & GOLDBERG, A. C. 2011. Autoimmune hepatitis, HLA and extended haplotypes. *Autoimmun Rev,* 10**,** 189-93.

OROZCO, G., BARTON, A., EYRE, S., DING, B., WORTHINGTON, J., KE, X. & THOMSON, W. 2011. HLA-DPB1-COL11A2 and three additional xMHC loci are independently associated with RA in a UK cohort. *Genes Immun,* 12**,** 169-75.

OROZCO, G., EERLIGH, P., SANCHEZ, E., ZHERNAKOVA, S., ROEP, B. O., GONZALEZ-GAY, M. A., LOPEZ-NEVOT, M. A., CALLEJAS, J. L., HIDALGO, C., PASCUAL-SALCEDO, D., BALSA, A., GONZALEZ-ESCRIBANO, M. F., KOELEMAN, B. P. & MARTIN, J. 2005. Analysis of a functional BTNL2 polymorphism in type 1 diabetes, rheumatoid arthritis, and systemic lupus erythematosus. *Hum Immunol,* 66**,** 1235-41.

PAI, A. A., PRITCHARD, J. K. & GILAD, Y. 2015. The genetic and mechanistic basis for variation in gene regulation. *PLoS Genet,* 11**,** e1004857.

PAN, Q., NING, Y., CHEN, L. Z., ZHANG, S., LIU, Z. Z., YANG, X. X., WEI, W., WEI, H., LI, Q. G., YUE, H. N. & WANG, J. X. 2012. Association of MHC class-III gene polymorphisms with ER-positive breast cancer in Chinese Han population. *Genet Mol Res,* 11**,** 4299-306.

PAN, Q., SHAI, O., LEE, L. J., FREY, B. J. & BLENCOWE, B. J. 2008. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet,* 40**,** 1413-5.

PARSONS, K. T., KWOK, W. W., GAUR, L. K. & NEPOM, G. T. 2000. Increased frequency of HLA class II alleles DRB1*0301 and DQB1*0201 in Lambert-Eaton myasthenic syndrome without associated cancer. *Hum Immunol,* 61**,** 828-33.

PARTS, L., LIU, Y. C., TEKKEDIL, M. M., STEINMETZ, L. M., CAUDY, A. A., FRASER, A. G., BOONE, C., ANDREWS, B. J. & ROSEBROCK, A. P. 2014. Heritability and genetic basis of protein level variation in an outbred population. *Genome Res,* 24**,** 1363-70.

PATHAN, S., GOWDY, R. E., COONEY, R., BECKLY, J. B., HANCOCK, L., GUO, C., BARRETT, J. C., MORRIS, A. & JEWELL, D. P. 2009. Confirmation of the novel association at the BTNL2 locus with ulcerative colitis. *Tissue Antigens,* 74**,** 322-9.

PITTONI, V. & VALESINI, G. 2002. The clearance of apoptotic cells: implications for autoimmunity. *Autoimmun Rev,* 1**,** 154-61.

PLANT, K., FAIRFAX, B. P., MAKINO, S., VANDIEDONCK, C., RADHAKRISHNAN, J. & KNIGHT, J. C. 2014. Fine mapping genetic determinants of the highly variably expressed MHC gene ZFP57. *Eur J Hum Genet,* 22**,** 568-71.

PREBLE, O. T., BLACK, R. J., FRIEDMAN, R. M., KLIPPEL, J. H. & VILCEK, J. 1982. Systemic lupus erythematosus: presence in human serum of an unusual acid-labile leukocyte interferon. *Science,* 216**,** 429-31.

PURDOM, E., SIMPSON, K. M., ROBINSON, M. D., CONBOY, J. G., LAPUK, A. V. & SPEED, T. P. 2008. FIRMA: a method for detection of alternative splicing from exon array data. *Bioinformatics,* 24**,** 1707-1714.

RAUCH, I., HAINZL, E., ROSEBROCK, F., HEIDER, S., SCHWAB, C., BERRY, D., STOIBER, D., WAGNER, M., SCHLEPER, C., LOY, A., URICH, T., MULLER, M., STROBL, B., KENNER, L. & DECKER, T. 2014. Type I interferons have opposing effects during the emergence and recovery phases of colitis. *Eur J Immunol,* 44**,** 2749-60.

RAVENSCROFT, J. C., SURI, M., RICE, G. I., SZYNKIEWICZ, M. & CROW, Y. J. 2011. Autosomal dominant inheritance of a heterozygous mutation in SAMHD1 causing familial chilblain lupus. *Am J Med Genet A,* 155A**,** 235-7.

RELLE, M. & SCHWARTING, A. 2012. Role of MHC-linked susceptibility genes in the pathogenesis of human and murine lupus. *Clin Dev Immunol,* 2012**,** 584374.

RHODES, B., FURNROHR, B. G., ROBERTS, A. L., TZIRCOTIS, G., SCHETT, G., SPECTOR, T. D. & VYSE, T. J. 2012. The rs1143679 (R77H) lupus associated variant of ITGAM (CD11b) impairs complement receptor 3 mediated functions in human monocytes. *Ann Rheum Dis,* 71**,** 2028-34.

RHODES, D. A., STAMMERS, M., MALCHEREK, G., BECK, S. & TROWSDALE, J. 2001. The cluster of BTN genes in the extended major histocompatibility complex. *Genomics,* 71**,** 351-62.

ROBINSON, T., KARIUKI, S. N., FRANEK, B. S., KUMABE, M., KUMAR, A. A., BADARACCO, M., MIKOLAITIS, R. A., GUERRERO, G., UTSET, T. O., DREVLOW, B. E., ZAACKS, L. S., GROBER, J. S., COHEN, L. M., KIROU, K. A., CROW, M. K., JOLLY, M. & NIEWOLD, T. B. 2011. Autoimmune disease risk variant of IFIH1 is associated with increased sensitivity to IFN-alpha and serologic autoimmunity in lupus patients. *J Immunol,* 187**,** 1298-303.

RODRIGUEZ-CARRIO, J., LOPEZ, P. & SUAREZ, A. 2015. Type I IFNs as biomarkers in rheumatoid arthritis: towards disease profiling and personalized medicine. *Clin Sci (Lond),* 128**,** 449-64.

RONNBLOM, L. 2011. The type I interferon system in the etiopathogenesis of autoimmune diseases. *Ups J Med Sci,* 116**,** 227-37.

RONNBLOM, L. E., ALM, G. V. & OBERG, K. E. 1991. Autoimmunity after alpha-interferon therapy for malignant carcinoid tumors. *Ann Intern Med,* 115**,** 178-83.

ROTHER, K. I., BROWN, R. J., MORALES, M. M., WRIGHT, E., DUAN, Z., CAMPBELL, C., HARDIN, D. S., POPOVIC, J., MCEVOY, R. C., HARLAN, D. M., ORLANDER, P. R. & BROD, S. A. 2009. Effect of ingested interferon-alpha on beta-cell function in children with new-onset type 1 diabetes. *Diabetes Care,* 32**,** 1250-5.

RUIZ-NARVAEZ, E. A., FRASER, P. A., PALMER, J. R., CUPPLES, L. A., REICH, D., WANG, Y. A., RIOUX, J. D. & ROSENBERG, L. 2011. MHC region and risk of systemic lupus erythematosus in African American women. *Hum Genet,* 130**,** 807-15.

SALLOUM, R., FRANEK, B. S., KARIUKI, S. N., RHEE, L., MIKOLAITIS, R. A., JOLLY, M., UTSET, T. O. & NIEWOLD, T. B. 2010. Genetic variation at the IRF7/PHRF1 locus is associated with autoantibody profile and serum interferon-alpha activity in lupus patients. *Arthritis Rheum,* 62**,** 553-61.

SALOMONSEN, J., CHATTAWAY, J. A., CHAN, A. C., PARKER, A., HUGUET, S., MARSTON, D. A., ROGERS, S. L., WU, Z., SMITH, A. L., STAINES, K., BUTTER, C., RIEGERT, P., VAINIO, O., NIELSEN, L., KASPERS, B., GRIFFIN, D. K., YANG, F., ZOOROB, R.,

GUILLEMOT, F., AUFFRAY, C., BECK, S., SKJODT, K. & KAUFMAN, J. 2014. Sequence of a complete chicken BG haplotype shows dynamic expansion and contraction of two gene lineages with particular expression patterns. *PLoS Genet,* 10**,** e1004417.

SÁNCHEZ-PERNAUTE, O., PÉREZ-FERRO, M., GABUCIO, A. & BENÍTEZ, R. 2015. The Expression of Interferon- Induced Protein with Tetracopeptide Repeats 1 in Monocytes, a Shortcut to the Interferon Signature in Lupus. *BJ Autoimmun Disod.,* 1**,** 1.

SCOTT, A. P., LAING, N. G., MASTAGLIA, F., NEEDHAM, M., WALTER, M. C., DALAKAS, M. C. & ALLCOCK, R. J. 2011. Recombination mapping of the susceptibility region for sporadic inclusion body myositis within the major histocompatibility complex. *J Neuroimmunol,* 235**,** 77-83.

SEKINE, H., FERREIRA, R. C., PAN-HAMMARSTROM, Q., GRAHAM, R. R., ZIEMBA, B., DE VRIES, S. S., LIU, J., HIPPEN, K., KOEUTH, T., ORTMANN, W., IWAHORI, A., ELLIOTT, M. K., OFFER, S., SKON, C., DU, L., NOVITZKE, J., LEE, A. T., ZHAO, N., TOMPKINS, J. D., ALTSHULER, D., GREGERSEN, P. K., CUNNINGHAM-RUNDLES, C., HARRIS, R. S., HER, C., NELSON, D. L., HAMMARSTROM, L., GILKESON, G. S. & BEHRENS, T. W. 2007. Role for Msh5 in the regulation of Ig class switch recombination. *Proc Natl Acad Sci U S A,* 104**,** 7193-8.

SESTAK, A. L., NATH, S. K., SAWALHA, A. H. & HARLEY, J. B. 2007. Current status of lupus genetics. *Arthritis Res Ther,* 9**,** 210.

SHABALIN, A. A. 2012. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics,* 28**,** 1353-8.

SHIMANE, K., KOCHI, Y., SUZUKI, A., OKADA, Y., ISHII, T., HORITA, T., SAITO, K., OKAMOTO, A., NISHIMOTO, N., MYOUZEN, K., KUBO, M., HIRAKATA, M., SUMIDA, T., TAKASAKI, Y., YAMADA, R., NAKAMURA, Y., KAMATANI, N. & YAMAMOTO, K. 2013. An association analysis of HLA-DRB1 with systemic lupus erythematosus and rheumatoid arthritis in a Japanese population: effects of *09:01 allele on disease phenotypes. *Rheumatology (Oxford),* 52**,** 1172-82.

SIGURDSSON, S., NORDMARK, G., GÖRING, H.H., LINDROOS, K., WIMAN, A.C., STURFELT, G., JÖNSEN, A., RANTAPÄÄ-DAHLQVIST, S., MÖLLER, B., KERE, J., KOSKENMIES, S., WIDÉN, E., ELORANTA, M.L., JULKUNEN, H., KRISTJANSDOTTIR, H., STEINSSON, K., ALM, G., RÖNNBLOM, L., SYVÄNEN, A.C. 2005. Polymorphisms in the tyrosine kinase 2 and interferon regulatory factor 5 genes are associated with systemic lupus erythematosus. *Am J Hum Genet*, 76,528-37.

SIRIKONG, M., TSUCHIYA, N., CHANDANAYINGYONG, D., BEJRACHANDRA, S., SUTHIPINITTHARM, P., LUANGTRAKOOL, K., SRINAK, D., THONGPRADIT, R., SIRIBOONRIT, U. & TOKUNAGA, K. 2002. Association of HLA-DRB1*1502-DQB1*0501 haplotype with susceptibility to systemic lupus erythematosus in Thais. *Tissue Antigens,* 59**,** 113-7.

SIROTA, M., SCHAUB, M. A., BATZOGLOU, S., ROBINSON, W. H. & BUTTE, A. J. 2009. Autoimmune disease classification by inverse association with SNP alleles. *PLoS Genet,* 5**,** e1000792.

SMIKLE, M., CHRISTIAN, N., DECEULAER, K., BARTON, E., ROYE-GREEN, K., DOWE, G., ANDERSON, N. & NICHOLSON, G. 2002. HLA-DRB alleles and systemic lupus erythematosus in Jamaicans. *South Med J,* 95**,** 717-9.

SMITH, J. K., CHI, D. S., KRISHNASWAMY, G., SRIKANTH, S., REYNOLDS, S. & BERK, S. L. 1996. Effect of interferon alpha on HLA-DR expression by human buccal epithelial cells. *Arch Immunol Ther Exp (Warsz),* 44**,** 83-8.

SPAGNOLO, P., SATO, H., GRUTTERS, J. C., RENZONI, E. A., MARSHALL, S. E., RUVEN, H. J., WELLS, A. U., TZOUVELEKIS, A., VAN MOORSEL, C. H., VAN DEN BOSCH, J. M., DU BOIS, R. M. & WELSH, K. I. 2007. Analysis of BTNL2 genetic polymorphisms in British and Dutch patients with sarcoidosis. *Tissue Antigens,* 70**,** 219-27.

SPECTOR, T. D. & WILLIAMS, F. M. 2006. The UK Adult Twin Registry (TwinsUK). *Twin Res Hum Genet,* 9**,** 899-906.

SRINIVASAN, K., SHIUE, L., HAYES, J. D., CENTERS, R., FITZWATER, S., LOEWEN, R., EDMONDSON, L. R., BRYANT, J., SMITH, M., ROMMELFANGER, C., WELCH, V., CLARK, T. A., SUGNET, C. W., HOWE, K. J., MANDEL-GUTFREUND, Y. & ARES, M. 2005. Detection and measurement of alternative splicing using splicing-sensitive microarrays. *Methods,* 37**,** 345-359.

STEEL, C. M., LUDLAM, C. A., BEATSON, D., PEUTHERER, J. F., CUTHBERT, R. J., SIMMONDS, P., MORRISON, H. & JONES, M. 1988. HLA haplotype A1 B8 DR3 as a risk factor for HIV-related disease. *Lancet,* 1**,** 1185-8.

STRANGER, B. E., FORREST, M. S., DUNNING, M., INGLE, C. E., BEAZLEY, C., THORNE, N., REDON, R., BIRD, C. P., DE GRASSI, A., LEE, C., TYLER-SMITH, C., CARTER, N., SCHERER, S. W., TAVARE, S., DELOUKAS, P., HURLES, M. E. & DERMITZAKIS, E. T. 2007. Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science,* 315**,** 848-53.

SUGGS, M. J., MAJITHIA, V., LEWIS, R. E. & CRUSE, J. M. 2011. HLA DRB1*1503 allelic haplotype predominance and associated immunodysregulation in systemic lupus erythematosus. *Exp Mol Pathol,* 91**,** 548-62.

SWANSON, C. L., WILSON, T. J., STRAUCH, P., COLONNA, M., PELANDA, R. & TORRES, R. M. 2010. Type I IFN enhances follicular B cell contribution to the T cell-independent antibody response. *J Exp Med,* 207**,** 1485-500.

TADA, M., ISHII-WATABE, A., MAEKAWA, K., FUKUSHIMA-UESAKA, H., KUROSE, K., SUZUKI, T., KANIWA, N., SAWADA, J., KAWASAKI, N., NAKAJIMA, T. E., KATO, K., YAMADA, Y., SHIMADA, Y., YOSHIDA, T., URA, T., SAITO, M., MURO, K., DOI, T., FUSE, N., YOSHINO, T., OHTSU, A., SAIJO, N., OKUDA, H., HAMAGUCHI, T., SAITO, Y. & MATSUMURA, Y. 2012. Genetic polymorphisms of FCGR2A encoding Fcgamma receptor IIa in a Japanese population and functional analysis of the L273P variant. *Immunogenetics,* 64**,** 869-77.

TAKAHASHI, M., YASUNAMI, M., KUBOTA, S., TAMAI, H. & KIMURA, A. 2006. HLA-DPB1*0202 is associated with a predictor of good prognosis of Graves' disease in the Japanese. *Hum Immunol,* 67**,** 47-52.

TAKAHASHI, N., NANIWA, T. & BANNO, S. 2008. Successful use of etanercept in the treatment of acute lupus hemophagocytic syndrome. *Mod Rheumatol,* 18**,** 72-5.

TANAKA, N., SATO, M., LAMPHIER, M. S., NOZAWA, H., ODA, E., NOGUCHI, S., SCHREIBER, R. D., TSUJIMOTO, Y. & TANIGUCHI, T. 1998. Type I interferons are essential mediators of apoptotic death in virally infected cells. *Genes Cells,* 3**,** 29-37.

TANG, Y., LUO, X., CUI, H., NI, X., YUAN, M., GUO, Y., HUANG, X., ZHOU, H., DE VRIES, N., TAK, P. P., CHEN, S. & SHEN, N. 2009. MicroRNA-146A contributes to abnormal activation of the type I interferon pathway in human lupus by targeting the key signaling proteins. *Arthritis Rheum,* 60**,** 1065-75.

THEOFILOPOULOS, A. N., BACCALA, R., BEUTLER, B. & KONO, D. H. 2005. Type I interferons (alpha/beta) in immunity and autoimmunity. *Annu Rev Immunol,* 23**,** 307-36.

TIAN, Y., LIAO, I. H., ZHAN, X., GUNTHER, J. R., ANDER, B. P., LIU, D., LIT, L., JICKLING, G. C., CORBETT, B. A., BOS-VENEMAN, N. G., HOEKSTRA, P. J. & SHARP, F. R. 2011. Exon expression and alternatively spliced genes in Tourette Syndrome. *Am J Med Genet B Neuropsychiatr Genet,* 156B**,** 72-8.

TRAUGOTT, U. & LEBON, P. 1988. Multiple sclerosis: involvement of interferons in lesion pathogenesis. *Ann Neurol,* 24**,** 243-51.

TRINCHIERI, G. 2010. Type I interferon: friend or foe? *J Exp Med,* 207**,** 2053-63.

TSAO, B. P. 2004. Update on human systemic lupus erythematosus genetics. *Curr Opin Rheumatol,* 16**,** 513-21.

TSOKOS, G. C. 2011. Systemic lupus erythematosus. *N Engl J Med,* 365**,** 2110-21.

TURNBULL, A. K., KITCHEN, R. R., LARIONOV, A. A., RENSHAW, L., DIXON, J. M. & SIMS, A. H. 2012. Direct integration of intensity-level data from Affymetrix and Illumina microarrays improves statistical power for robust reanalysis. *BMC Med Genomics,* 5**,** 35.

UPPAL, S. S., HAYAT, S. J. & RAGHUPATHY, R. 2009. Efficacy and safety of infliximab in active SLE: a pilot study. *Lupus,* 18**,** 690-7.

URIBE, A. G., MCGWIN, G., JR., REVEILLE, J. D. & ALARCON, G. S. 2004. What have we learned from a 10-year experience with the LUMINA (Lupus in Minorities; Nature vs. nurture) cohort? Where are we heading? *Autoimmun Rev,* 3**,** 321-9.

VANDIEDONCK, C., TAYLOR, M. S., LOCKSTONE, H. E., PLANT, K., TAYLOR, J. M., DURRANT, C., BROXHOLME, J., FAIRFAX, B. P. & KNIGHT, J. C. 2011. Pervasive haplotypic variation in the spliceo-transcriptome of the human major histocompatibility complex. *Genome Res,* 21**,** 1042-54.

VAUGHN, S. E., KOTTYAN, L. C., MUNROE, M. E. & HARLEY, J. B. 2012. Genetic susceptibility to lupus: the biological basis of genetic risk found in B cell signaling pathways. *J Leukoc Biol,* 92**,** 577-91.

VAVASSORI, S., KUMAR, A., WAN, G. S., RAMANJANEYULU, G. S., CAVALLARI, M., EL DAKER, S., BEDDOE, T., THEODOSSIS, A., WILLIAMS, N. K., GOSTICK, E., PRICE, D. A., SOUDAMINI, D. U., VOON, K. K., OLIVO, M., ROSSJOHN, J., MORI, L. & DE LIBERO, G. 2013. Butyrophilin 3A1 binds phosphorylated antigens and stimulates human gammadelta T cells. *Nat Immunol,* 14**,** 908-16.

VIGANO, A., PINTI, M., NASI, M., MORETTI, L., BALLI, F., MUSSINI, C., BRICALLI, D., SALA, N., BUGARINI, R., VELLA, S., PRINCIPI, N. & COSSARIZZA, A. 2001. Markers of cell death-activation in lymphocytes of vertically HIV-infected children naive to highly active antiretroviral therapy: the role of age. *J Allergy Clin Immunol,* 108**,** 439-45.

VIGNAL, C., BANSAL, A. T., BALDING, D. J., BINKS, M. H., DICKSON, M. C., MONTGOMERY, D. S. & WILSON, A. G. 2009. Genetic association of the major histocompatibility complex with rheumatoid arthritis implicates two non-DRB1 loci. *Arthritis Rheum,* 60**,** 53-62.

VIKEN, M. K., BLOMHOFF, A., OLSSON, M., AKSELSEN, H. E., POCIOT, F., NERUP, J., KOCKUM, I., CAMBON-THOMSEN, A., THORSBY, E., UNDLIEN, D. E. & LIE, B. A. 2009. Reproducible association with type 1 diabetes in the extended class I region of the major histocompatibility complex. *Genes Immun,* 10**,** 323-33.

VOISIN, S., ALMEN, M. S., ZHELEZNYAKOVA, G. Y., LUNDBERG, L., ZAREI, S., CASTILLO, S., ERIKSSON, F. E., NILSSON, E. K., BLUHER, M., BOTTCHER, Y., KOVACS, P., KLOVINS, J., RASK-ANDERSEN, M. & SCHIOTH, H. B. 2015. Many obesity-associated SNPs strongly associate with DNA methylation changes at proximal promoters and enhancers. *Genome Med,* 7**,** 103.

VOLANAKIS, J. E., ZHU, Z. B., SCHAFFER, F. M., MACON, K. J., PALERMOS, J., BARGER, B. O., GO, R., CAMPBELL, R. D., SCHROEDER, H. W., JR. & COOPER, M. D. 1992. Major histocompatibility complex class III genes and susceptibility to immunoglobulin A deficiency and common variable immunodeficiency. *J Clin Invest,* 89**,** 1914-22.

WALL, P. K., LEEBENS-MACK, J., CHANDERBALI, A. S., BARAKAT, A., WOLCOTT, E., LIANG, H., LANDHERR, L., TOMSHO, L. P., HU, Y., CARLSON, J. E., MA, H., SCHUSTER, S. C., SOLTIS, D. E., SOLTIS, P. S., ALTMAN, N. & DEPAMPHILIS, C. W. 2009. Comparison of next generation sequencing technologies for transcriptome characterization. *BMC Genomics,* 10**,** 347.

WANG, Y., EWART, D., CRABTREE, J. N., YAMAMOTO, A., BAECHLER, E. C., FAZELI, P. & PETERSON, E. J. 2015. PTPN22 Variant R620W Is Associated With Reduced Toll-like Receptor 7-Induced Type I Interferon in Systemic Lupus Erythematosus. *Arthritis Rheumatol,* 67**,** 2403-14.

WATANABE, N. & NAKAJIMA, H. 2012. Coinhibitory molecules in autoimmune diseases. *Clin Dev Immunol,* 2012**,** 269756.

WATERS, H., KONRAD, P. & WALFORD, R. L. 1971. The distribution of HL-A histocompatibility factors and genes in patients with systemic lupus erythematosus. *Tissue Antigens,* 1**,** 68-73.

WESTRA, H. J., ARENDS, D., ESKO, T., PETERS, M. J., SCHURMANN, C., SCHRAMM, K., KETTUNEN, J., YAGHOOTKAR, H., FAIRFAX, B. P., ANDIAPPAN, A. K., LI, Y., FU, J., KARJALAINEN, J., PLATTEEL, M., VISSCHEDIJK, M., WEERSMA, R. K., KASELA, S., MILANI, L., TSEREL, L., PETERSON, P., REINMAA, E., HOFMAN, A., UITTERLINDEN, A. G., RIVADENEIRA, F., HOMUTH, G., PETERSMANN, A., LORBEER, R., PROKISCH, H., MEITINGER, T., HERDER, C., RODEN, M., GRALLERT, H., RIPATTI, S., PEROLA, M., WOOD, A. R., MELZER, D., FERRUCCI, L., SINGLETON, A. B., HERNANDEZ, D. G., KNIGHT, J. C., MELCHIOTTI, R., LEE, B., POIDINGER, M., ZOLEZZI, F., LARBI, A., WANG DE, Y., VAN DEN BERG, L. H., VELDINK, J. H., ROTZSCHKE, O., MAKINO, S., SALOMAA, V., STRAUCH, K., VOLKER, U., VAN MEURS, J. B., METSPALU, A., WIJMENGA, C., JANSEN, R. C. & FRANKE, L. 2015. Cell Specific eQTL Analysis without Sorting Cells. *PLoS Genet,* 11**,** e1005223.

WESTRA, H. J. & FRANKE, L. 2014. From genome to function by studying eQTLs. *Biochim Biophys Acta,* 1842**,** 1896-1902.

WESTRA, H. J., PETERS, M. J., ESKO, T., YAGHOOTKAR, H., SCHURMANN, C., KETTUNEN, J., CHRISTIANSEN, M. W., FAIRFAX, B. P., SCHRAMM, K., POWELL, J. E., ZHERNAKOVA, A., ZHERNAKOVA, D. V., VELDINK, J. H., VAN DEN BERG, L. H., KARJALAINEN, J., WITHOFF, S., UITTERLINDEN, A. G., HOFMAN, A., RIVADENEIRA, F., T HOEN, P. A., REINMAA, E., FISCHER, K., NELIS, M., MILANI, L., MELZER, D., FERRUCCI, L., SINGLETON, A. B., HERNANDEZ, D. G., NALLS, M. A., HOMUTH, G., NAUCK, M., RADKE, D., VOLKER, U., PEROLA, M., SALOMAA, V., BRODY, J., SUCHY-DICEY, A., GHARIB, S. A., ENQUOBAHRIE, D. A., LUMLEY, T., MONTGOMERY, G. W., MAKINO, S., PROKISCH, H., HERDER, C., RODEN, M., GRALLERT, H., MEITINGER, T., STRAUCH, K., LI, Y., JANSEN, R. C., VISSCHER, P. M., KNIGHT, J. C., PSATY, B. M., RIPATTI, S., TEUMER, A., FRAYLING, T. M., METSPALU, A., VAN MEURS, J. B. & FRANKE, L. 2013. Systematic identification of

trans eQTLs as putative drivers of known disease associations. *Nat Genet,* 45**,** 1238-43.

WILLCOCKS, L. C., LYONS, P. A., CLATWORTHY, M. R., ROBINSON, J. I., YANG, W., NEWLAND, S. A., PLAGNOL, V., MCGOVERN, N. N., CONDLIFFE, A. M., CHILVERS, E. R., ADU, D., JOLLY, E. C., WATTS, R., LAU, Y. L., MORGAN, A. W., NASH, G. & SMITH, K. G. 2008. Copy number of FCGR3B, which is associated with systemic lupus erythematosus, correlates with protein expression and immune complex uptake. *J Exp Med,* 205**,** 1573-82.

WILSON, A. G., CLAY, F. E., CRANE, A. M., CORK, M. J. & DUFF, G. W. 1995. Comparative genetic association of human leukocyte antigen class II and tumor necrosis factor-alpha with dermatitis herpetiformis. *J Invest Dermatol,* 104**,** 856-8.

WILSON, A. G., DE VRIES, N., POCIOT, F., DI GIOVINE, F. S., VAN DER PUTTE, L. B. & DUFF, G. W. 1993. An allelic polymorphism within the human tumor necrosis factor alpha promoter region is strongly associated with HLA A1, B8, and DR3 alleles. *J Exp Med,* 177**,** 557-60.

WU, J., WILSON, J., HE, J., XIANG, L., SCHUR, P. H. & MOUNTZ, J. D. 1996. Fas ligand mutation in a patient with systemic lupus erythematosus and lymphoproliferative disease. *J Clin Invest,* 98**,** 1107-13.

WU, X. M., WANG, C., ZHANG, K. N., LIN, A. Y., KIRA, J., HU, G. Z., QU, X. H., XIONG, Y. Q., CAO, W. F. & GONG, L. Y. 2009a. Association of susceptibility to multiple sclerosis in Southern Han Chinese with HLA-DRB1, -DPB1 alleles and DRB1-DPB1 haplotypes: distinct from other populations. *Mult Scler,* 15**,** 1422-30.

WU, Y. L., HAUPTMANN, G., VIGUIER, M. & YU, C. Y. 2009b. Molecular basis of complete complement C4 deficiency in two North-African families with systemic lupus erythematosus. *Genes Immun,* 10**,** 433-45.

WUCHERPFENNIG, K. W. & SETHI, D. 2011. T cell receptor recognition of self and foreign antigens in the induction of autoimmunity. *Semin Immunol,* 23**,** 84-91.

YANG, W., SHEN, N., YE, D. Q., LIU, Q., ZHANG, Y., QIAN, X. X., HIRANKARN, N., YING, D., PAN, H. F., MOK, C. C., CHAN, T. M., WONG, R. W., LEE, K. W., MOK, M. Y., WONG, S. N., LEUNG, A. M., LI, X. P., AVIHINGSANON, Y., WONG, C. M., LEE, T. L., HO, M. H., LEE, P. P., CHANG, Y. K., LI, P. H., LI, R. J., ZHANG, L., WONG, W. H., NG, I. O., LAU, C. S., SHAM, P. C., LAU, Y. L. & ASIAN LUPUS GENETICS, C. 2010. Genome-wide association study in Asian populations identifies variants in ETS1 and WDFY4 associated with systemic lupus erythematosus. *PLoS Genet,* 6**,** e1000841.

YANG, W., TANG, H., ZHANG, Y., TANG, X., ZHANG, J., SUN, L., YANG, J., CUI, Y., ZHANG, L., HIRANKARN, N., CHENG, H., PAN, H. F., GAO, J., LEE, T. L., SHENG, Y., LAU, C. S., LI, Y., CHAN, T. M., YIN, X., YING, D., LU, Q., LEUNG, A. M., ZUO, X., CHEN, X., TONG, K. L., ZHOU, F., DIAO, Q., TSE, N. K., XIE, H., MOK, C. C., HAO, F., WONG, S. N., SHI, B., LEE, K. W., HUI, Y., HO, M. H., LIANG, B., LEE, P. P., CUI, H., GUO, Q., CHUNG, B. H., PU, X., LIU, Q., ZHANG, X., ZHANG, C., CHONG, C. Y., FANG, H., WONG, R. W., SUN, Y., MOK, M. Y., LI, X. P., AVIHINGSANON, Y., ZHAI, Z., RIANTHAVORN, P., DEEKAJORNDEJ, T., SUPHAPEETIPORN, K., GAO, F., SHOTELERSUK, V., KANG, X., YING, S. K., ZHANG, L., WONG, W. H., ZHU, D., FUNG, S. K., ZENG, F., LAI, W. M., WONG, C. M., NG, I. O., GARCIA-BARCELO, M. M., CHERNY, S. S., SHEN, N., TAM, P. K., SHAM, P. C., YE, D. Q., YANG, S., ZHANG, X. & LAU, Y. L. 2013. Meta-analysis followed by replication identifies loci in or near CDKN1B, TET3, CD80, DRAM1, and ARID5B as associated with systemic lupus erythematosus in Asians. *Am J Hum Genet,* 92**,** 41-51.

YANG, Y., CHUNG, E. K., ZHOU, B., LHOTTA, K., HEBERT, L. A., BIRMINGHAM, D. J., ROVIN, B. H. & YU, C. Y. 2004. The intricate role of complement component C4 in human systemic lupus erythematosus. *Curr Dir Autoimmun,* 7**,** 98-132.

YASUTOMO, K., HORIUCHI, T., KAGAMI, S., TSUKAMOTO, H., HASHIMURA, C., URUSHIHARA, M. & KURODA, Y. 2001. Mutation of DNASE1 in people with systemic lupus erythematosus. *Nat Genet,* 28**,** 313-4.

YOSHIDA, T., KATO, K., HORIBE, H., OGURI, M., FUKUDA, M., SATOH, K., AOYAGI, Y., SHINKAI, S., NOZAWA, Y. & YAMADA, Y. 2011. Association of a genetic variant of BTN2A1 with chronic kidney disease in Japanese individuals. *Nephrology (Carlton),* 16**,** 642-8.

YUAN, J., ZHAO, D., WU, L., XU, X., PANG, Y., ZHANG, J., MA, Y., LIU, J. & WANG, J. 2015. FCGR3B copy number loss rather than gain is a risk factor for systemic lupus erythematous and lupus nephritis: a meta-analysis. *Int J Rheum Dis,* 18**,** 392-7.

ZELLER, T., WILD, P., SZYMCZAK, S., ROTIVAL, M., SCHILLERT, A., CASTAGNE, R., MAOUCHE, S., GERMAIN, M., LACKNER, K., ROSSMANN, H., ELEFTHERIADIS, M.,

SINNING, C. R., SCHNABEL, R. B., LUBOS, E., MENNERICH, D., RUST, W., PERRET, C., PROUST, C., NICAUD, V., LOSCALZO, J., HUBNER, N., TREGOUET, D., MUNZEL, T., ZIEGLER, A., TIRET, L., BLANKENBERG, S. & CAMBIEN, F. 2010. Genetics and beyond--the transcriptome of human monocytes and disease susceptibility. *PLoS One,* 5**,** e10693.

ZHANG, Z., SONG, L., MAURER, K., PETRI, M. A. & SULLIVAN, K. E. 2010. Global H4 acetylation analysis by ChIP-chip in systemic lupus erythematosus monocytes. *Genes Immun,* 11**,** 124-33.

ZHOU, M., LI, L. H., PENG, H., LI, R., FENG, C. C., XU, W. D., LENG, R. X., PAN, H. F. & YE, D. Q. 2014. Decreased ITGAM and FcgammaRIIIA mRNA expression levels in peripheral blood mononuclear cells from patients with systemic lupus erythematosus. *Clin Exp Med,* 14**,** 269-74.

ZHOU, X., LEE, J. E., ARNETT, F. C., XIONG, M., PARK, M. Y., YOO, Y. K., SHIN, E. S., REVEILLE, J. D., MAYES, M. D., KIM, J. H., SONG, R., CHOI, J. Y., PARK, J. A., LEE, Y. J., LEE, E. Y., SONG, Y. W. & LEE, E. B. 2009. HLA-DPB1 and DPB2 are genetic loci for systemic sclerosis: a genome-wide association study in Koreans with replication in North Americans. *Arthritis Rheum,* 60**,** 3807-14.

ZHOU, Y., WU, J., KUCIK, D. F., WHITE, N. B., REDDEN, D. T., SZALAI, A. J., BULLARD, D. C. & EDBERG, J. C. 2013. Multiple lupus-associated ITGAM variants alter Mac-1 functions on neutrophils. *Arthritis Rheum,* 65**,** 2907-16.